

Where should GATT go next?

The GATT negotiations now mercifully reaching the end of their six-year course are not arcane procedures invented to keep diplomats busy, but a vital means by which the benefits of technology are widely shared.

THE six years so far spent on the negotiation of a supplement to the General Agreement on Tariffs and Trade (GATT) may not, after all, be wasted. Two weeks ago, it could easily have been different. That was when Mr Raymond McSharry, the Irish member of the European Commission who has spent most of the past six months arguing with agricultural interests in the United States about European subsidies of vegetable oil, threw in his hand, saying that the president of the commission, M. Jacques Delors, should take personal responsibility for the negotiations — and for the failure that then seemed certain.

Luckily, wiser councils have prevailed. McSharry (not Delors) is back at the negotiating table. Although the row over oilseeds, involving Europe and the United States bilaterally, is peripheral to the GATT agreement proper, the United States has made it clear that it will not sign the second until there is agreement on the first. That prospect, and the publication of a list of European products on which the United States plans to levy tariffs of 200 per cent two weeks from now, seems to have frightened governments into their senses. Nothing is certain. France continues to object to the oilseeds deal on offer. But a new version of GATT by the end of February seems once again more likely than unlikely.

Not surprisingly. Failure would calamitously exaggerate the economic recession from which the whole world (except China and some parts of South-East Asia) is now suffering. (Failure would also inevitably encourage governments to forsake multilateral arrangements for the bilateral but less efficient deals they instinctively prefer.) Indeed, it is even possible that only a new GATT agreement can bring the recession to an end. The Organisation for Economic Cooperation and Development estimates that the new agreement will eventually add \$200 billion to the annual value of world trade — a pie big enough to square a good many obdurate economic circles. It is in nobody's interest to hazard a prize of that magnitude.

But it is also now self-evident that the past six years of negotiations (due to have ended in 1990) are the wrong way to liberalize world trade. GATT is a treaty to which most industrialized and many developing countries belong, but which no communist or ex-communist countries have yet joined. Its central principles are that countries may not discriminate among trading partners, keeping out imports from one but accepting them from another, and may not unfairly subsidize their own industries or flood international markets with goods at unfair low prices. (To police the system, GATT runs a quasijudicial arbitration procedure which, among many

other things, has found Europe's oilseed subsidies to be unfair.) Then, on the principle that restraints on trade (by import tariffs or otherwise) impoverish everybody, GATT organizes negotiations (such as the present) in which members strike deals among themselves on the levels at which tariffs apply to various classes of traded goods and the ranges of goods to which GATT's rules apply. Agriculture is now contentious because its products have not previously been included.

But GATT does not embody objective measures of progress. Negotiations usually depend on members' willingness to trade off seemingly equitable reductions of the amounts of import tax they will collect if tariffs are reduced. But there is no correspondence between the tax collected by a tariff and its economic impact. Increasing the US tariff on European white wine (threatened for December) by 200 per cent, for example, would have the paradoxical effect that the United States collects no tax from that source; European imports will dry up, and Americans will drink Californian or Australian wine instead (but they should not overlook Upper New York State).

Conversely, a country (say Nigeria) that does not manufacture computer hardware does itself a disservice by levying tariffs on imported equipment. True, it collects some tax, but why should it thereby hinder people's use of techniques that could help to make the country rich? When countries are entirely dependent on outside sources for particular kinds of goods, it is an expensive folly to levy tariffs on them. (Relying on tariffs for tax revenues is usually damaging.) More generally, governments appear to have been lulled (perhaps by the existence of GATT) into the belief that only negotiations can deliver the benefits of free trade; unilateral decisions to abolish tariffs, without reciprocal action by trading partners, is the simplest way if the balance of payments can take the strain. But tariffs (or export subsidies in the case of European oilseeds) are not the only or even the most serious impediments to free trade.

That is why, if the present round of tariff cuts is successfully negotiated, GATT should follow a different tack in whatever comes next. The past six years have shown how easily principles are buried by detail. Adam Smith's explanation of the "wealth of nations" was that a modern state functions most efficiently, and thus has more of life's goods to share, when individuals specialize at the tasks they do best — cobblers make shoes, farmers grow grain and then satisfy each others' needs for grain and shoes respectively by trading with each other. GATT admirably aims to extrapolate the same principle



onto the international stage. It cannot hope to be successful while its members' misgivings about free trade are either overlooked or, worse, are allowed themselves to become absolute restraints.

The most immediate anxiety concerns jobs. A country abolishing import tariffs on, say, mousetraps exposes its own mousetrap manufacturers to greater competition, under which they may wilt — and those who work for them may lose their jobs. The benefits (cheaper or even better mousetraps in the shops) may be outweighed by the social disruption that follows. So it should be the norm and not the exception that programmes to reduce the mousetrap tariff are phased over intervals of time comparable with the economic lifetime of mousetrap factories and the time taken to retrain mousetrap workers at other tasks — ideally at tasks adding greater value. Decade-long transitional periods should be the general rule, but GATT should be able to insist that transitional arrangements are not indefinitely extended. One of the scandals in world trade is the Multi-Fibre Agreement (in which GATT played no part) by means of which rich countries have restrained textile imports from poor countries for the best part of two decades. (The bilateral agreement between the United States and Japan has the makings of a similar scandal.)

Chauvinism is another impediment. Steelmaking may no longer be a source of national pride, but most governments believe they should demonstrate their patriotism by operating airlines, television broadcasting stations, telecommunications networks and the like. Even when these enterprises are not publicly owned, there are irksome and even meaningless restrictions on foreign ownership (as with airlines in the United States), often based on flimsy assertions of strategic national interest. That is why one of GATT's next objectives should be free trade in commercial companies. The subject has at least been opened up during the present negotiations, which originally set out to liberalize international trade in services (life insurance, for example).

The long-term need is for a clear distinction between the substantial and the insubstantial aspects of chauvinism. Governments subsidizing national airlines, for example, impoverish their own people (and, by insisting on reciprocity with foreign airlines, air travellers in general). That is folly. But the same governments have a right to insist on whatever standards of safety they choose. Similarly, governments may reasonably require that foreign providers of services should satisfy local rules of financial probity, that foreign employers of local labour should follow local practice if people have to be dismissed and so on.

Ambiguities in the protection of intellectual property are another serious impediment to the improvement of world trade. They have been given some attention in the past six years, but not enough. Rich governments have been pressing, with some success, for relief from the piracy of patents on drugs, semiconductor devices and other innovations. There are two difficulties: patent protection usually applies only nationally, while many developing countries have asserted the right to use others' patents for their domestic purposes. The second is part of the long-standing argument about the

transfer of technology from rich to poor countries.

Is it too much to ask that GATT should take up the cause of devising a genuinely international regime for patent protection? The sheer cost of the present arrangements, under which inventors must seek protection in every country in which they think they need it, is in itself a waste of resources. UNESCO in the past has tilted at this windmill, but that does not mean the cause is lost. But would developing countries now making free with other people's patents ever concede inventors' title? They might if that were traded for the abolition of the present restraints on poor countries' potential exports — the Multi-Fibre Agreement is merely the most notorious of them.

The best recipe for GATT's future is thus a more imaginative one. Negotiating an international patents regime would not be child's play. Abandoning rich countries' restrictive quotas on poor countries' exports would similarly require heart-searching and far-sightedness by the well-off states. But the outcome would be a much bigger step towards Adam Smith's beneficent division of labour than even what may emerge from the present round of negotiations. But a valuable side-effect would be the erosion of the cynicism by which foreign assistance is at present directed towards the improvement of the economies of developing countries, which are then told that their potential exports are not wanted. GATT has been a success so far. It could be a still bigger one. □

Cleaner gasoline

Oxygenated gasoline for US cars is meant to improve air quality in 38 cities.

For years while European nations have squarely faced the high cost of carbon with hefty taxes on gasoline for cars and trucks, the United States has, for political reasons, refused to levy a tax that would send a clear signal that gasoline is an expensive (and polluting) commodity that must be conserved.

Finally, in response to the Clean Air Act of 1990, the US government has required 38 cities with polluted air to sell cleaner but more expensive oxygenated gasoline during the winter months when concentrations of carbon monoxide rise because combustion engines work less efficiently in the cold. The oxygenated gasoline, the only kind that will be allowed in many areas, is expected to reduce carbon monoxide emissions by 15–20 per cent over three years.

Although the United States continues to resist higher gasoline taxes (President-elect Bill Clinton may change that), the result of mandatory consumption of oxygenated gas is the equivalent of a tax increase of as much as 10 cents a gallon, first because the cleaner gas costs 3–5 cents a gallon more and second because there is a small loss of fuel economy. In the Washington DC area, for instance, the cost of a gallon of gas will probably rise from an average of \$1.14 to \$1.24 a gallon.

This change in policy comes none too soon and should be a harbinger of more sensible (that is more extensive) gas taxing policies in the future. □

UK science budget to grow despite economic woes and cap on salaries

London. The British government's announcement last week that it plans to increase the science budget by just under 4 per cent next year dispelled rumours circulating among the heads of the country's research councils that the economic crisis would deal a serious blow to science. Government ministers have labelled the increase, slightly higher than the anticipated rate of inflation, as a "modest rise" in spending on research.

"The settlement is a good deal better than we had been led to believe", says Sir Mark Richmond, chairman of the Science and Engineering Research Council (SERC). Similarly Dai Rees, secretary of the Medical Research Council, put out a statement saying that the council is "pleased with today's news".

There was widespread disappointment, however, that the government failed to keep to last year's promise to secure a real growth in the science budget and that government scientists — including those working in universities — will receive increases of no more than 1.5 per cent, a ceiling being imposed on all public sector workers over the next 12 months.

"Scientists' salaries — hardly over-generous at the moment — will fall even further in comparison with earnings in industry and abroad", says Rees. "If we are to work to improve the health of the nation, we cannot afford to put ourselves at a disadvantage which, however short-term in intention, will be long-term in effect."

In its announcement, the government said that the science budget, which includes the work of the five research councils, will be £1.164 billion (US\$1.85 billion) for the fiscal year starting next April. After deducting £125 million being transferred from the universities' budget under the dual support system, this represents an increase of 3.69 per cent over this year's figure of £1.050 billion (minus £48 million in transfers) and a real increase of 1 per cent. But it is still £17 million less than the government last year provisionally agreed to spend next year.

Provisional figures for subsequent years have also been reduced, to £1.231 billion in 1994–95 (a reduction of £40 million) and £1.265 billion the following year. When inflation is taken into account, the result is a flat research budget for the next three years.

Government officials have been keen to put a positive slant on the figures. William Waldegrave, minister of science, said at a press conference that he was "reasonably pleased" with the outcome of what he described as "the toughest expenditure rounds that I have been involved in over the past decade. . . . In a very difficult year, it shows

that spending on science remains one of the government's highest priorities."

Detailed figures are not yet available for research spending by other government departments. Waldegrave said, however, that "from a preliminary look, it appears as if their R&D spending will on average be

money out of its own budget.

A spokesman for the Economic and Social Research Council said that next year's funding constraints mean that "all flexibility has now virtually disappeared. We are facing real cuts in our budget, which in particular means the loss of planned research programmes on issues like risk, human behaviour and training for employment."

The decision to limit pay increases provides short-term relief for the science budget. Salaries account for about 40 per cent of this budget, and the move will save about £3 million.

Waldegrave believes that most scientists would prefer to have the government cap their salaries than reduce the amount of research it funds. But others have warned that any advantage could be short-lived. A spokesman for the Association of University Teachers described the settlement as "particularly bad news" in the light of the career uncertainties facing British scientists.

"Much of the impact [of the pay restrictions] will depend on how long the situation lasts", says Richmond. "If people think that it is going to last for two or three years, then they are going to start looking for jobs outside research."

Not surprisingly, the government's decision to allocate less money to science next year than it had previously promised has produced sharp criticism from the pressure group Save British Science. "Everyone welcomed the first sign of a real increase which the government promised last year. Now that has been wiped out," says John Mulvey, professor of physics at the University of Oxford and secretary of the SBS group. "We cannot express any pleasure at that, even though we understand that hard decisions have to be taken during a recession."

SBS is pinning its hopes on the contents of a white paper (policy document) on the organization of British science, due out early next year. "We hope that, at the very least, the White Paper will contain some targets to be aimed at over the next four or five years that will see the science base properly funded and help secure the desperately needed investment", says Mulvey.

Waldegrave refused to speculate on its contents, but he suggested that the creation of a new Office of Science and Technology (OST), separating responsibility for science from the Department of Education and Science, shows the government's commitment to research. "I believe that if the OST had not existed, the budget would have been worse", he said.

David Dickson



"The settlement is a good deal better than we had been led to believe."

Sir Mark Richmond

about comparable with ours — in other words, they will be about level in real terms."

The Advisory Board for the Research Councils, which must recommend to Waldegrave how this money should be divided, faces some difficult decisions. The Science and Engineering Research Council, for example, faces a major increase in the costs of its international subscriptions, in particular to the European Laboratory for Particle Physics (CERN) and the European Space Agency (ESA), because of the recent devaluation of the pound.

"The situation is a good deal worse than it was a year ago", says Richmond, pointing out that the council must spend about £10 million more next year on international subscriptions. During the previous currency crisis a few years ago, the government eventually agreed to cover the shortfall; this time SERC has been told to find the

Yeltsin's ecology adviser chips away at a sorry past

Washington. Alexei Yablokov knows that he's pushing a stone uphill as counsellor to the Russian president, Boris Yeltsin, for ecology and public health. A zoologist by training and the former director of the Institute of Developmental Biology in Moscow, Yablokov now spends a good deal of his time uncovering new evidence of the catastrophic environmental legacy of the former Soviet government. But he knows that economic considerations are paramount, and that his message must be tempered by harsh fiscal realities.

A case in point is the recent decision to continue operating a family of Russian nuclear reactors of the Chernobyl type on the grounds that the country's need for electric power outweighs the dangers posed by the graphite RBMK reactors. Yablokov believes that the country would be better off if it abandoned nuclear power and switched to gas turbines, a move made easier by his estimate that Russia will reduce its energy consumption by 30 per cent in converting from a military to a civilian economy. But his views failed to carry the day at a recent cabinet meeting. "Only the minister of the environment and public health supported me", he says. "All the other ministers are against me for economic reasons."

Yablokov left his laboratory in 1989 for politics because of the opportunity to influence the country's environmental policies, and he does not regret his decision. "As counselor to the president I help to arrange his schedule, influencing where he goes

and what he does", he says.

Last month, Yablokov arranged for Yeltsin to visit the Volga River, where he signed an order protecting the sturgeon, whose well-being is a source of national pride. And earlier this year, Yeltsin proposed legislation to add an environmental component to efforts to redefine national security following the end of the Cold War. In both cases, Yablokov used his position as resident expert to strengthen Yeltsin's policies on environmental preservation. Yablokov also successfully appealed against a decision to build a dam on the Katun River in eastern Siberia, persuading Yeltsin to delay the project pending further study of its effect on the environment.

Environmentalists elsewhere applaud his work. "He has been very influential in shaping government decision-making in Russia on environmental issues", says Ronald Kendall of Clemson University and president of the 3,000-member Society of Environmental Toxicology and Chemistry, which gave Yablokov its first-ever science award during its annual convention last week in Cincinnati, Ohio. "He's kept the public and the president aware of past, present and future environmental problems from radioactivity to DDT", says Murray Feshbach of Georgetown University, a demographer and longtime Kremlinologist.

Although Yablokov does not direct a large bureaucracy and has fewer than a dozen people on his personal staff, Kendall says that he is influential in redirecting

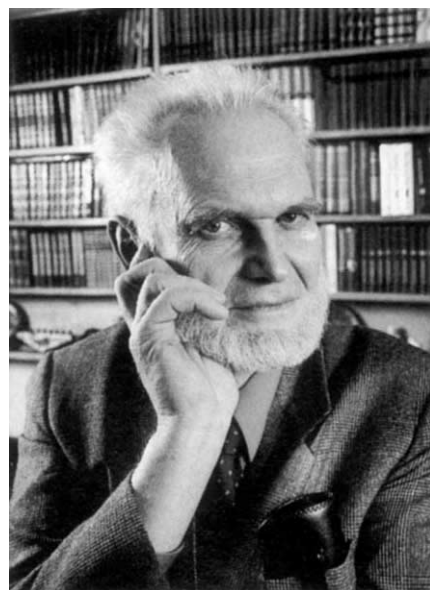


Photo by David Huxford

Yablokov has president's ear.

funding among ministries. Yablokov has also acted as host to the US vice president-elect, Al Gore, during a visit as a US senator, and he can be expected to exploit that relationship as Russia searches for ways to finance costly cleanups.

Apart from the economy, Yablokov must also contend with a paucity of useful data. Speaking last week at the US Armed Forces Radiobiology Research Institute in Washington, Yablokov shook his head at a question about followup studies on those affected by past environmental disasters. "Whenever we analyse past data, we have to realize that we have no right to believe it", he replied. "It was collected for political purposes, and often it does not tell us what we need to know."

Jeffrey Mervis

Russian academy institutes told to set priorities

Moscow. The Russian Academy of Sciences is losing control over all but its weakest institutes as its entire research structure adapts to the need for greater self-sufficiency.

The various branches of the academy have been ordered to rank their research programmes since the ruling presidium decided last month to sharply reduce its portfolio. Those not seen as priorities will be expected to find other sources of support or will be forced to close. The abolition of entire institutes is another possibility.

The presidium's action is not expected to affect institutes that have spent the past two years preparing to stand on their own, in particular the institutes of general physics, molecular genetics and molecular biology. The director of the Institute of Chemical Physics, Vitaly Goldanskyy, expects no reduction in the size of his staff, while the Institute of General Physics regards a

decision to reduce its staff by 10 per cent as simply a pruning of those who have been the least productive. Other institutes expect to supplement their reduced state subsidy by leasing part of their buildings to commercial enterprises or by obtaining grants from international organizations.

A new trade union of employees has criticized the presidium's actions and disputes its estimate that the academy faces a budget deficit of more than 1 billion rubles. It argues that self-sufficiency should not be the sole factor in deciding what research to support and that the differing costs of infrastructure must also be considered. "A rigid selection of priorities at the expense of all else will result in a system in which only the strongest survive", says Vitaly Sobolev, chairman of the trade union, which has asked the presidium to reconsider.

A greater threat to many institutes is the

delay in implementing an emergency measure that would make it easier for researchers to benefit from outside support. The measure, signed in February by the Russian president, Boris Yeltsin, would give scientists the right to open accounts in foreign banks and exempt from taxes of as much as 40 per cent purchases of equipment and sums of hard currency. It would also prohibit the state from dictating the ways in which an outside grant could be used.

"I am afraid that it will remain a draft forever", says Igor Nikolayev of the Ministry of Science. "I don't see any other way out than to call on the scientists themselves to lobby this document at any level of power accessible to them." Without such changes, Russian researchers may have little to gain from efforts by Western and European countries to shore up their laboratories.

Vladimir Pokrowsky

Livermore laboratory buys Russian papers on laser technology

Washington. Lawrence Livermore National Laboratory in California is buying \$300,000 worth of unclassified research reports and hardware from 19 Russian laboratories and one Ukrainian institute in the latest example of scientists in the former Soviet Union finding novel, capitalist ways to support their research.

The research is unclassified and not related to work on weapons. It involves laser technologies with potential applications in monitoring the atmosphere for improved weather prediction and pollution control and chemical and optical technologies for breaking down hazardous wastes.

The Russian and Ukrainian scientists are expected to deliver the material within three months. Although some of the ex-Soviet scientists may be invited to Livermore to answer questions, none will do research at the US laboratory nor work jointly with their US colleagues.

"It's a straight business deal", says a spokesman for the laboratory. "We get technical material, in English, and hardware, and they get US dollars plus the professional satisfaction of knowing that their work is being put to use."

The purchase was negotiated by a four-man team that visited Russia this past summer. It is separate from an agreement among the United States, the European Communities and Japan to support international research centres in Moscow and Kiev employing former Soviet weapons scientists on projects of interest to their industrialized partners. But that initiative, scheduled to be launched during the summer, is mired in government red tape and has yet to make its first award.

It is also different from the increasingly common arrangement by which teams of Russian scientists are hired by a Western company to carry out joint research. Last month, for example, AT&T Bell Laboratories announced a contract with the A.F. Ioffe Physico-Technical Institute in St Petersburg in which 27 Ioffe scientists will do basic research in semiconductor physics and semiconductor lasers. In February, the laboratory's material science division signed a similar contract with some 90 scientists at the General Physics Institute in Moscow.

Unlike the Livermore purchases, the contract between AT&T and the Russian institutes provides money for salaries, equipment and travel. Although Western companies have been accused of simply hiring cheap labour, the contracts are intended to keep research teams intact in the face of declining support by the governments of Russia and the other former Soviet republics.

Jeffrey Mervis

Costs curtail but do not kill European space laboratory

Granada, Spain. Disputes between Germany and France over their favoured man-in-space projects were patched up at last week's ministerial meeting in Granada of the European Space Agency (ESA) held to decide ESA's plans to the end of the century. But the compromises failed to resolve the fundamental issue of how to pay the bill for the prestigious space laboratory project, designed as a component of the international space station Freedom.

ESA's determination to present a united public face averted a looming crisis similar

in the DM2.5 billion project from 14 per cent to 10 per cent and suggested a review of its overall cost, raising the prospect of a delay that would be unacceptable to the Germans. At the same time, Spain announced that it was cutting by half its 6 per cent contribution, leaving the project only 90 per cent funded.

Despite its strong interest, Germany refused to increase its share. As a temporary solution, the ministers agreed to continue development while ESA seeks a reduction of 5 per cent over the next three years in total



The ESA ministers may have been more friendly in public than in private.

to that at last year's meeting in Munich, when ministers from ESA's 13 members overturned long-held plans because they could no longer afford the spiralling costs. This time, ESA council members have been given a year to trim 15 per cent from their original proposals costing DM52.8 billion (US\$35 billion), but disagreement between the major players, France and Germany, continues to pose problems. Just before the meeting began, for example, the council failed to reach a consensus on funding for the space laboratory, known as the APM (attached pressurized module).

The major sticking point has been Europe's two major contributions to the international space station mission — the APM and the space plane Hermes. France's elaborate hopes for Hermes, intended to ferry astronauts to the space station, have been thwarted by German efforts to keep costs 'realistic' (see *Nature* 357, 351; 1992). France had agreed to pay 40 per cent of the project's costs, which under ESA rules entitles it to 40 per cent of the ensuing industrial contracts. But the project has been downgraded at best to an unmanned vehicle that may not stand on its own.

Germany had taken a similarly large stake in APM, which will be attached to Freedom as an experimental area for astronauts. Last week, France said it would reduce its stake

costs. Funding for the remaining 5 per cent was left unresolved.

ESA was given until March next year to make good that shortfall. Its president, Hubert Curien, the French science minister, said that countries interested in the project will be asked to increase their stake. Italy, whose 25 per cent interest in the project makes it the second largest investor, has informally indicated its willingness to step into the breach. Spain may soften its stance if the ESA council next month agrees to adjust the general contributions from members to account for currency devaluations.

Less controversial was a decision to seek wider international cooperation, particularly in the development of space vehicles with Russia (see *Nature* 359, 176; 1992). The resolution also asks for "foundations to be laid for cooperation with Japan" and the "deepening" of established cooperation with the United States.

Projects for Earth observations won strong support from delegations eager to embrace environmental protection. In switching its financial interest from the APM to its associated multisatellite system, the 'polar platform', France now overtakes Britain as the major player, with an estimated 23.6 per cent share, but British Aerospace is expected to retain its position as prime contractor.

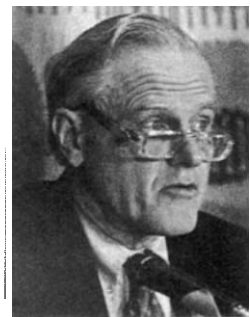
Alison Abbott

Petition warns of pending global environmental crisis

Washington. More than 1,500 prominent scientists from around the world have endorsed a "Warning to Humanity" that says far-reaching political and economic changes are needed to avert an environmental catastrophe. The petition, sponsored by the Union of Concerned Scientists (UCS) and unveiled on 18 November in Washington DC,

takes aim at a host of environmental problems from global warming to overpopulation.

Although such petitions are common, UCS says that its four-page document is distinguished by the unusually large number of signers, among



Henry Kendall

them 98 Nobel prizewinners. It calls for a more efficient use of resources and an end to activities that damage the environment. Its goals, including eliminating poverty and giving "women control over their own reproductive decisions", would be funded in part with money now spent on military defence. The petition is being sent to government leaders around the world.

Supporters, including dozens from the developing world and many who seldom lend their names to causes outside their profession, say they were attracted by the broad scope of the petition and its vague but ambitious recommendations, which range from eliminating poverty to ensuring sexual equality. The even-handed tone of the petition, which blames both developing and developed nations for environmental damage, also appealed to many.

Some of the petition's statements are open to scientific debate. For example, it declares that "acid [rain] precipitation is causing widespread injury" and that ozone depletion is a "critical stress" to the environment. The petition acknowledges that the effects of global warming are still unclear but says that "uncertainty...cannot excuse complacency or delay in facing the threats".

Henry Kendall, a Nobel prizewinner in physics and chairman of the UCS, says that the organizers gathered the endorsements of scientists from different fields "to verify the aggregate pieces" of the petition. Edward O. Wilson, a Crafoord prize-winning biologist at Harvard, says it would have been "poor citizenship" to say that "I don't speak to things outside my field." There comes a time when scientists have to assume at least minimal responsibility." **Traci Watson**

Biotechnology soared and fell in a roller-coaster 12 months

Washington. A report by stock market analysts on the past year in US biotechnology reflects the mixed fortunes of an industry in which most companies are still losing money despite improving financial indicators. The availability and cost of new capital and an unpredictable regulatory environment continue to be a concern.

The report* points out that in the year ending June 1992 the industry raised more than US\$3 billion in public equity, exceeding the total for the entire decade of the 1980s. In January, Amgen Inc. of Thousand Oaks, California, became the first biotechnology company to be included in the *Fortune* 500 business list, with sales of \$645 million in 1991, largely from its first two products, Epogen and Neupogen.

But other companies were not so fortunate. In April, Centocor Inc. of Malvern, Pennsylvania, failed to win approval from the US Food and Drug Administration (FDA) for the company's main product, Centoxin (a human monoclonal antibody for the treatment of Gram-negative sepsis), and its stock price plummeted. It was also a year in which, despite an increase of 29 per cent in revenues for public companies (to \$4.5 billion), four out of five publicly held companies lost money and do not expect to make a profit before the middle of the decade.

Increasingly, companies are trying to accelerate commercialization and spread their risk. By acquiring licences for late-stage technologies from other companies and by exchanging products, companies are generating revenues in the short-term and improving their chances of survival. Through acquisitions, partnerships and strategic alliances, they are acquiring manufacturing capabilities and sales forces to produce and distribute their own lines.

The report's coauthor, G. Steven Burrill of Ernst & Young, found that strategic alliances are occurring earlier, often before a company goes public and sometimes even before it tries to obtain venture capital. These alliances add value and credibility to a company before it goes public.

The past year saw little change in the demographics of the industry; smaller-sized companies still predominate, with 76 per cent of companies having 50 or fewer employees. There have been double-digit percentage increases in both the number of companies and numbers of employees, to 1,231 and 79,000 respectively. As in the past few years, when the industry is broken down by market sector, the survey found that more companies are specializing in human health care than all other market segments combined.

Although there has been little change in

the geographical landscape, with the San Francisco, New York and Boston areas having the largest concentration of companies, Burrill says that there has been considerable activity in the Seattle and San Diego areas and some around Washington DC.

Agricultural biotechnology has in the past been the poor relation of the human health care segment. Apart from the difficulties of genetically modifying organisms, indifference by investors and continued uncertainty surrounding its regulation have been barriers to commercialization.

However, two recent developments on the regulatory front could pave the way for agricultural biotechnology. In May, the FDA announced that foods developed using genetic engineering pose no new or special safety risks to the consumer and should be subject to the same standards of regulation as other foods (see *Nature* 357, 352; 1992). And earlier this month, the US Department of Agriculture announced plans to ease regulation of US field tests of genetically engineered crops (see *Nature* 360, 94; 1992). The report adds that optimism must be tempered by a need to inform the public about the safety and benefits of genetically engineered agricultural products.

The industry's record-setting performance in the public equity markets came during a time when the United States was in the grips of a recession. Investor interest was also piqued by the huge success of Amgen, whose stock price tripled in 1991. The romance with Wall Street lasted 18 months, to June 1992, during which time the industry raised more than \$5 billion in public equity and 67 of the now 225 public companies completed initial public offerings.

All that changed when Centocor's leading product veered off the regulatory track and the need to contain health care costs became an issue in the 1992 presidential election. Investors turned away from health care stocks and by June the financing window had slammed shut.

In the few months since then, there has been renewed interest in the biotechnology industry by venture capital firms, and Burrill predicts that 1992 could be a record year for start-up companies. The biotechnology industry is "awash with venture capital", he says, having raised \$1.6 billion during the first six months of this year compared with \$1.3 billion for the whole of 1991. Burrill expects a similar amount of new venture capital to be invested in the second half of the year and the number of new start-up companies to exceed 80. **Diane Gershon**

* *Biotech 93 Accelerating Commercialization*: Ernst & Young's Seventh Annual Report on the Biotech Industry (Ernst & Young, San Francisco, California, 1992).

Plant collections endangered in Eastern Europe and Russia

Paris. More than half a million varieties of 2,500 plant species stored in gene banks in Eastern Europe and the former Soviet Union are at risk, according to a report issued last month by the Food and Agriculture Organisation and the International Board for Plant Genetic Resources. Immediate

pean collections will disappear because most are held in long-term storage facilities. But the harsh economic conditions still pose a serious threat: for example, at the Institute for Plant Genetic Resources in Sadovo, Bulgaria, with tens of thousands of accessions, compressors in one of the three cold rooms are broken and those in a second room are unreliable. If the remaining cold room ceases to function, the institute does not have hard currency to buy the needed Russian spare parts.

In addition, most of the collections do not have funding commitments beyond the end of the year. With the institutes now obliged to produce and to sell seed to pay for essential staff and operating expenses, many are terminating support for gene banks. The idea that such facilities should be supported by the government has not been embraced by the

governments of Eastern Europe.

Although the situation appears grim, a relatively small infusion of cash could reap large dividends. Emile Frison, group leader in Europe for the plant genetics resources board, estimates that "\$2.5 million over four to five years" would protect a multi-billion-dollar investment. He says that such a rescue package might buy enough time to persuade the relevant governments that germplasm collections require public funds.

Some money is forthcoming. The US Agriculture Department has agreed to spend \$90,000 in the next three years to develop computing facilities at the Vavilov Institute. Although the gene banks have been carefully compiled and monitored, little of the information is on computers and available for dissemination. The entire St Petersburg collection, for example, is contained in one set of handwritten catalogues. The US Agency for International Development has promised \$400,000 contingent on contributions from other countries, and the United Nations Development Programme has committed as much as \$100,000. The World Bank has asked the plant genetics board to coordinate the flow of aid to these collections to make sure that the money goes to those most in need.

European countries are broadly supportive of preserving the collections but it is too soon to say how much money, if any, will be forthcoming. Norway, for example, has suggested that any aid should be part of an effort to rescue germplasm of global significance rather than one intended primarily to preserve a research collection. **Declan Butler**

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Dr Karl Hammer, Head of the German Genebank Gatersleben

The Vavilov Institute in St Petersburg

international assistance is needed to avoid the damage that could be caused from a depressed economy and a series of decisions to privatize plant breeding institutes without providing continued public funding for the gene banks associated with them.

The Soviets have collections unmatched elsewhere because of the work of Nikolai Vavilov, the first president of the Academy of Agricultural Sciences, beginning in the 1920s. But with funding for agricultural research programmes being reduced by as much as 80 per cent since the breakup of the Soviet Union, his emphasis on plant genetic resources is being abandoned.

Gene banks in the former Soviet Union are most at risk because of the lack of long-term storage facilities, at -20°C , for nearly 350,000 accessions at the Vavilov Institute in St Petersburg and its 17 experimental stations. More than half the collection is kept in paper bags at room temperature in St Petersburg, and another 120,000 items are stored at $5-8^{\circ}\text{C}$ at Kuban. These accessions must be regrown, multiplied and dried every five years, a process that is labour intensive, and there are insufficient funds for equipment to extend their life.

The fate of a quarter of the Soviet collection, mostly subtropical species, is even more uncertain. These stations are owned by the newly independent states but are poorly equipped compared with gene banks at Kuban and St Petersburg. Moreover, the new governments are unlikely to have either the resources or the interest to maintain this germplasm.

There is less risk that the Eastern Euro-

CHINA IN BRIEF

Beijing. Construction could begin as early as 1994 on China's controversial Three Gorges Dam approved earlier this year (see *Nature* **356**, 736; 1992). The project is being designed in three parts — the dam and related facilities, a plan for the resettlement of more than one million people and a network to transmit electricity to eastern and central China — and Yang Zhen-Huai, water conservancy minister, said last month that the overall design for the dam is almost completed and will be unveiled next year. The government has already built a dedicated railway line to the remote and mountainous site and will, by the end of the year, have spent more than 3 billion yuan (US\$550 million) on new homes for those to be relocated. Opposition to the project persists, but the Chinese press has been unwilling to publicize their concerns.

■ In a major relaxation of its decade-long grip on graduate education, the Chinese State Education Commission will allow institutions of higher education to enrol students whose fees are paid by industry or who are self-supporting. Selected institutions will also be allowed to use state research money to fund graduate students. Under the present system, which has awarded 21,000 master's degrees and 7,000 doctoral degrees since 1981, the commission selected each student and, upon graduation, assigned them to a job. The new rules will make universities more responsive to the demand for highly educated personnel by allowing other enterprises to sponsor students or having students pay their own way and be free to find work after they graduate. The government now supports 80,000 graduate students a year, a figure expected to shrink under the new system.

■ China's first institution to train foreigners in traditional medicine and acupuncture opened last month in Tianjin, the country's third largest city. The International College of Chinese Traditional Medicine hopes eventually to enrol 500 foreign students, divided among undergraduate, graduate and short-term advanced study students. The college is authorized by the government to grant master's degrees in eight specialities and a doctoral degree in acupuncture.

■ The Chinese Academy of Sciences may elect members from outside the country under a new constitution approved earlier this year. Internationally renowned scientists who have made important contributions to the country's scientific and technological progress will be considered in future elections, to be held every two years. Another change on the horizon is the creation of a free-standing academy of engineering and technology sciences.

You Qin Li

New contract with DOE laboratories guarantees academic freedoms

Washington. Researchers at some of the largest US national laboratories are about to get their first guarantee of something their academic colleagues take for granted: scientific and intellectual freedom. As part of a proposed new management contract with the University of California (UC), the Los Alamos, Lawrence Livermore and Lawrence Berkeley national laboratories will for the first time explicitly assure their 17,600 combined employees of the same rights as UC researchers.

These rights, although always professed, have never been in the contract with the US Department of Energy (DOE) to manage the three laboratories, two of which began as nuclear weapons laboratories and still maintain large classified research programmes. This omission has not been academic: in the past decade, dozens of scientists have been threatened and even muzzled for criticizing everything from the Star Wars missile defence system to Desert Storm and the US fusion programmes. Others have been forbidden to attend conferences attended by researchers from the Soviet Union and other nations seen as adversaries.

One who has suffered greatly from the blurring of scientific freedom is Hugh Dewitt, a theoretical physicist at Lawrence Livermore. After criticizing the laboratory's work on a X-ray laser weapon for the Star Wars project, Dewitt suffered the ignominy of monthly performance assessments and poor reviews and eventually left the laboratory for a year. By battling back on a platform of academic freedom, Dewitt was

able to clear his name. But without the explicit guarantee of such freedom, Dewitt fought for years to have his right to free speech recognized.

He and other laboratory researchers say that the new language — assuring laboratory scientists of the “equivalent rights and obligations of university faculty” in publication, participation in open debate and attendance at scientific and professional meetings — should put an end to those sort of struggles. “Next time I get in trouble, I’m certainly going to wave that paragraph around”, says Dewitt.

The proposed five-year contracts, released in final draft form last week, are the result of nearly a year and half of negotiations between DOE and UC. They must still be signed by DOE as well as by the UC board of regents, which is expected to approve them at a meeting this week.

But even getting this far has been an ordeal for the two parties. In 1989, UC faculty voted not to renew the contract because of concern over the university's affiliation with such a visible aspect of the US weapons complex. Although the regents subsequently voted to renew the contracts, DOE began having reservations of its own.

Beyond a lingering suspicion that the UC faculty would continue to prove an obstacle, DOE officials were concerned about UC's refusal to accept liability for environmental problems at the laboratories. Officials feared that could place DOE in the position of having to take the blame for — and swallow the cost of — management

problems that resulted in pollution.

Negotiations on the liability issue forced the UC regents to extend the current contract by two months in September. Yet when the dust settled, little had changed. UC will continue to be protected against liability or loss arising from its management of the laboratories. In return, UC accepted liability for any unallowable indirect costs, criminal fines and penalties or wilful misconduct by laboratory officials. DOE will put \$14 million annually into a contingency fund to pay for any such penalties. If UC can avoid such expenses, it can use the money as it sees fit, including supporting research involving UC and laboratory scientists.

Although the contract remains based on the idea of nonprofit management, UC's fee will more than double. In addition to its \$14 million contingency fee, UC will receive \$5 million a year for the ground on which the Lawrence Berkeley Laboratory stands (although it has always been UC property, the university has never before charged rent) and another \$5 million for a new supervisory office. That office, with at least 30 employees, will watch for the kind of abuses that have led to faculty protests in the past.

Overall, the university's annual management fee will increase from \$13 million this year to \$30 million. Although the state will end up with some of that money, much of it is expected to support new collaborative research between UC and laboratory scientists and more frequent transfers of laboratory technology to the outside world.

Christopher Anderson

Safeguards added for appeals in misconduct cases

Washington. Responding to complaints that federal misconduct investigations do not allow scientists enough opportunity to defend themselves, the US Public Health Service has agreed to offer a full court-like hearing to scientists found guilty of misconduct. But the new procedures, which could cost scientists thousands of dollars, may still not satisfy those who say that the system now deprives scientists of their right to a fair trial.

The Office of Research Integrity (ORI) is hoping the move will bring to misconduct investigations such basic rights as the ability to be represented by a lawyer, to cross-examine witnesses and to see evidence. This is intended to ameliorate concerns about its announced plans last month to limit further a scientist's rights under the Privacy Act to see his or her own records by obtaining an exemption to that act for misconduct inves-

tigations (see *Nature* **359**, 765; 1992). Although ORI says it will still seek such an exemption, it hopes that the new procedures will placate its critics.

Under the new regulations, any scientist found guilty of scientific misconduct after an investigation by ORI or the scientist's institution will automatically be offered an administrative hearing to review the case before either an administrative law judge or a panel of officials and scientists. Researchers will be able to question any ORI witnesses and evidence and to present their own evidence and witnesses in an effort to have the charges dropped. In a notice published in the *Federal Register* of 6 November, ORI said the new procedures would take effect immediately.

In the past, scientists facing debarment from future federal grants could request such a hearing, but they constitute only a

quarter of those found guilty of misconduct. The others face lesser penalties, such as loss of single grant, prohibition from sitting on federal review committees and official reprimands. Now all will be entitled to an appeal hearing.

However, it is not known how many will request such an appeal. Only three of 13 scientists facing debarment have done so, according to Lyle Bivens, the ORI director, and only about half of those accused of misconduct maintain their innocence throughout the investigation and, presumably, would be inclined to request a hearing after its conclusion.

Even for those, cost may be a factor. The hearing can take weeks, and legal fees can exceed \$50,000. Given that no scientists have yet won such a hearing, it may take a brave researcher indeed to risk so much for a day in court.

Christopher Anderson

Nature for non-English speakers

SIR — My advice to the Japanese journalist whose translation machine failed to explain the meaning of "Prides and prejudice" (*Nature* 359, 475; 1992), is to acquire the *Oxford English Dictionary*. The second edition, available in 20 volumes or on disk, gives as one of the definitions of "pride" "a group of lions forming a social unit". In addition, it lists the phrase "pride and prejudice", which is illustrated by 10 examples, the first dating from 1619, while the last gives the title of Jane Austen's novel, published in 1813. This makes an understanding of the phrase independent of one's knowledge of English literature.

Admittedly, anyone without access to sufficient funds in hard currency may have to make do with the two volumes of the *Shorter Oxford English Dictionary*. These do not include "pride and prejudice", but nevertheless give the necessary definitions of "pride", "Herculean", "Augean" and "Noah's Ark", although the entry on "Damascus" fails to inform on Paul's conversion.

All in all, baby language is hardly called for in an effort to make science intelligible. On the other hand, the use of slang would be totally out of place. Readers of *Nature* naturally need to understand its contents, but also use its style as a model in writing or giving talks in English. It is far less difficult to understand the meaning of a word in a foreign language than to acquire a feeling for its exact status. Many years ago, a German-speaking colleague, about to give a talk at a genetics meeting, asked me to look at the English of his manuscript. The only change needed was to substitute "hybrid" every time the word "bastard" occurred.

Equally, portmanteau adjectives, such as "DNA-dependent cytosol-controlled", owe nothing to the idiosyncrasies of English syntax and everything to scientists cutting corners in their use of language. If editors were to develop a healthy immune response in the face of these linguistic invaders, this would be a small step leading to a better understanding of science.

Ursula Mittwoch

Queen Mary & Westfield College,
Mile End Road, London E1 4NS, UK

SIR — Although I am neither a native speaker of English, nor among the happy few who understood the pun about "Prides and prejudice" at first glance (actually I learned the word for a group of lions from this cover), I hope you will maintain this kind of allusive writing. Critics should take into account that a headline of a News and Views item, Commentary and so on cannot actually

provide a short cut to spare one the trouble of reading the page by conveying the same information in a nutshell. At best it can attract the reader's attention so that s/he will read the article. Thus, a little bit of mystery in the headline need not be an impediment to understanding the body of the text.

Referring to your question of how to make an international journal more accessible to a polyglot audience, I plead for an educational approach. Instead of lowering the standard until the text becomes comprehensible to people without any knowledge of English (and unbearable for the rest of us), you should provide motivation for all scientists to reach a reasonable standard of English, including the cultural connotations of the language. One can certainly assume that the readers of *Nature* have intelligence and curiosity. Hence, a pun that they do not understand should be a challenge to think about the language. By keeping the 'literary' quality of its front pages, *Nature* will supply its readership with a valuable means to improve their English (for instance, if ever I meet a multitude of lions, I will always remember to address them as a pride and will never risk insulting them by confusing them with a pack of wolves, a flock of sheep or a gaggle of geese), whereas a process of receding to the smallest common denominator will accelerate the decline of scientific English towards a babytalk enriched by numbers and acronyms.

Michael Gross

Institut für Biophysik
and Physikalische Biochemie,
Universität Regensburg,
D-8400 Regensburg, Germany

Hunterian Institute

SIR — I read with great sadness Ian Mundell's article (*Nature* 358, 704; 1992) concerning the closure and removal of several research departments at the Hunterian Institute based at the Royal College of Surgeons of England, which effectively closes the Hunterian Institute one year before the 200th anniversary in 1993 of John Hunter's death. This is in direct conflict with the principles of John Hunter who stressed that true learning was possible only against a background of research. A well-known quotation of Hunter, directed to his friend Edward Jenner, was, "Why think? Why not do the experiment?" Sadly, that is not going to be possible now for many people at the Hunterian Institute.

Mundell said that although staff at the institute are unhappy about the changes, they are reluctant to speak publicly. As a

former PhD student and subsequently a research fellow at the Hunterian Institute, I feel it is the duty of all scientific staff to speak out strongly. For someone such as myself, just embarking on an independent scientific career, the loss of the Hunterian Institute is representative of a rapid demise in science in Great Britain and acts as another example to dissuade young graduates from entering science as a career (see *Nature* 355, 292; 1992). I left the Hunterian Institute earlier this year just before the news broke of its imminent closure. Like many other young scientists, including the final four PhD students who were trained in the department of biochemistry and cell biology, I am now working abroad. To some extent this is a normal career stage for postdoctoral graduates in order to gain a breadth of experience. But it is of great concern that there may be very few scientific jobs for which to return to Britain.

A factor that has compounded the closure of the institute is the lack of discussion between the staff, past council members and benefactors before the decision was made. During the final months of the institute, very little explanation was offered to members of staff as to why the institute was closing or what was to happen to their jobs. This lack of communication has resulted in some members of staff facing unemployment after many years of loyal and devoted service to the institute and college. The Royal College of Surgeons is a registered charity and relies upon public and industrial fund-raising for many of its activities; the appeal of the college to potential donors without many of its excellent research departments will be severely weakened. It is sad to think that a college of such excellence and once a great debating house in which the merits of Geoffroy's, Lamarck's and Cuvier's doctrines were discussed in such a radical and open way, no longer feels it necessary to continue that tradition. In the context of the National Health 'reforms' and publication of the Tomlinson report on the future of the London teaching hospitals, I hope a lesson will be learnt from this unhappy event; perhaps politicians and administrators of other institutes and medical schools will appreciate the excellent work their scientific colleagues are doing and attempt to help them in their endeavours, rather than closing them down. I feel that the Royal College of Surgeons, by closing down or moving most of the Hunterian Institute to make way for more office space, has done itself and British science a great disservice.

Paul Eggleton

Boston University School of Medicine,
80 East Concord Street, Boston,
Massachusetts 02118-2394, USA

Estimating Popper's impact

SIR — Skoyles¹, Bartley² and even Popper himself say that Popper's work has had little impact on professional philosophers. Skoyles¹ even claims that Popper has failed to influence sociologists and others studying science and society. The view that Popper's philosophy has been widely acknowledged only by scientists stems primarily from Bartley's account² of the development of the profession of philosophy of science and from his personal communications with Popper. However, judging from the references made to Popper in the *Science Citation Index (SCI)* and the *Social Sciences Citation Index (SSCI)*, this "failure" is grossly overstated.

According to the 1969–77 *SSCI*, comprising 800,000 source items and nearly 8 million references cited within them, Popper was the third most-cited author in the philosophy and history of science³, and his book *Logik der Forschung/The Logic of Scientific Discovery* (1935/1959) was the second most-cited book in this subject⁴. Of course, this does not directly measure Popper's influence on philosophers *per se* because the impact of his work could be on scholars in disciplines other than philosophy.

But when we compare the number of articles citing Popper in *SCI* (column A in the table), *SSCI* (excluding philosophy) (B) and *SSCI* (only philosophy) (C) during 1974–91, his impact in each field has been considerable.

Still, there is an apparently great difference between Popper's impact on the philosophy literature (C) as compared to the science (A) and social-sciences (B) literature. But this difference diminishes when we consider that the total number of ISI-indexed philosophy papers (50,006) is a small fraction of the total in the *SSCI* file (2,170,943) and the *SCI* file (9,741,144) during 1974–91.

Consider also that the annual average citations per cited author in the *SSCI* file has ranged from 3.4 to 4.4 during this time, and 7.1 to 8.8 in the *SCI*. In comparison, Popper's annual average citation rate in the *SSCI* was 234.9, and 79.0 in the *SCI*, during 1974–91. If we take into account only the philosophy

journals indexed in the *SSCI*, Popper's annual average citation rate is 39.6. So, despite Popper's modest estimation of himself, it is easy to agree with Bondi⁵ that "Popper's influence shines through".

B. I. B. Lindahl

Department of Geriatric Medicine,
Karolinska Institute,
S-141 86 Huddinge,
Sweden

A. Welljams-Dorof

Institute for Scientific Information,
3501 Market Street,
Philadelphia,
Pennsylvania 19104, USA

SIR — It is not surprising that there is no 'Popper' school¹. First, philosophers of science seem to be more concerned with the nature of reality and truth in general than the special features of science, and, second, Popper's contributions to the understanding of the nature of science seem to be overrated⁶. To his credit he has emphasized the creative aspects of science, but what is usually regarded as his most important contribution — that science proceeds by falsification — has severe limitations.

The basic idea that only falsification is important and one should not merely look for confirming instances was put forward by Claude Bernard in his book on experimental medicine in 1865⁷. But science works in a much more complex way, theories and experiments often being intimately linked. The history of science is filled with examples where scientists succeeded because they ignored falsification. Indeed, Gerald Holton⁸ has argued cogently that the graveyard of failed scientists is littered with those who did not practise a suspension of disbelief when their ideas were first shown to be wrong. As Francis Crick said: "A theory that fits all the facts is bound to be wrong, as some of the facts will be wrong".

The theory does not even resolve the problem of induction, for one wants to have a sufficient number of experimental falsifications to be persuaded. The idea also does not distinguish between science and nonscience because absurd ideas — such as that eating hamburgers will make you a good poet — are falsifiable. Worse still, Popper's philosophy is at heart relativistic, for there is no real measure of reliability of ideas or use of confirmations, and its emphasis on theory marginalizes empirical discovery.

Finally, consider applying the falsification test to his falsification hypothesis. How would one falsify it? If the evidence of history is not adequate what is? As Popper's champion, Medawar, pointed

out, science gets on very well without a philosophy of method.

Lewis Wolpert

Department of Anatomy
& Developmental Biology,
University College London,
Windeyer Building,
Cleveland Street, London W1P 6DB, UK

1. Skoyles, J. R. *Nature* **359**, 100 (1992).
2. Bartley, W. W. in *In Pursuit of Truth* (ed. Levinson, P.) 249–289 (Humanities, New Jersey, 1982).
3. Garfield, E. in *Essays of an Information Scientist* Vol. 3 (ed. Garfield, E.) 633–639 (ISI, Philadelphia, 1980).
4. Garfield, E. in *Essays of an Information Scientist* Vol. 3 (ed. Garfield, E.) 621–632 (ISI, Philadelphia, 1980).
5. Bondi, H. *Nature* **358**, 363 (1992).
6. Wolpert, L. *The Unnatural Nature of Science* (Faber, London, in the press).
7. Bernard, C. *Introduction to Experimental Medicine* (Macmillan, New York, 1927).
8. Holton, G. *The Scientific Imagination* (Cambridge University Press, 1978).

Women in science

SIR — Your leading article "Women in Science" (*Nature* **359**, 92; 1992) cannot go unchallenged. You argue that the proposal by the National Science Foundation to refuse financial support for scientific meetings unless women speakers are included in the programme is misguided, because it will indirectly lead to the conclusion "that women in science are not really as good as men". What nonsense! This conclusion will be reached only if the presentations by women over time are significantly inferior in quality to those of men.

There was no evidence for this at the exciting and stimulating Cold Spring Harbor Meeting on Mouse Molecular Genetics (26–30 August) where a high proportion of the speakers were women. Nor have I obtained this impression from a variety of other conferences in which women have been well represented. On the other hand, as a member of a National Institutes of Health Study Section I have repeatedly been disappointed to see proposals for meetings in which very few women have been invited. Invariably, the organizers of such meetings are well established men, while comparable meetings with women organizers include a reasonable proportion of articulate women doing good science. With time, some change is inevitable but, in my opinion, women have been waiting long enough and are tired of it. If faster change cannot be brought about by persuasion and feedback, as experience suggests, then stronger pressure has to be applied by funding agencies who want to promote a fair hearing for the women students and investigators they pay to train and support.

Brigid Hogan

Department of Cell Biology,
Vanderbilt University,
Nashville, Tennessee 37232, USA

NUMBER OF ARTICLES CITING WORKS BY
KARL POPPER

Years	A	B	C	Total
1974–76	111	447	109	667
1977–79	185	557	175	917
1980–82	240	648	101	989
1983–85	312	686	97	1095
1986–88	281	572	111	964
1989–91	293	607	119	1019
Total	1422	3517	712	5651

Conflicts of interest declared.

Referees are the most honourable men and women, but they should not fight shy of refusing to referee manuscripts in which their interest is overdirect.

REFEREES are a remarkable body of men and women on whom the whole process of peer review depends. It is a constant source of wonder that they are usually willing to make careful appraisals of other people's unpublished manuscripts in the most selfless fashion, carrying out a thorough analysis of what people working in their own field of interest have just done, making constructive comments both on the science described in a manuscript and on its presentation. Authors often agree that their intervention frequently improves the force of argument and the clarity of what is eventually published. That referees perform this service selflessly is a marvellous proof that science is indeed an altruistic profession.

Readers of what follows should bear that testimony to the importance of referees clearly in mind. Lapses from strict propriety are exceedingly rare and much less impressive than the sterling service regularly provided by dedicated people. That *Nature* has run into a crop of them in the past few months is probably a statistical fluctuation, not to be interpreted as proof of anything.

The first illustration that sticks in the mind goes back years. A referee had been sent a brief contribution on martensitic steel, on which he was the acknowledged expert in Britain. His reply caused some surprise, to say the least. On the content of the manuscript, the referee remarked that it was a worthy contribution to the field, but that its conclusions were not so exciting that the manuscript would sustain a claim on space in a journal such as *Nature*. That is a familiar comment, unremarkable in itself (but not a conclusive guide to action).

The surprise was what followed. The referee explained that he was taking the liberty of enclosing with his report a manuscript of his own, which he believed would make a very substantial contribution to understanding, and be entirely suitable for quick and wide publication. But casual inspection, confirmed by independent advice, suggested that the referee had little more to say than the original author.

The referee was not a crook. His paper was in no sense a copy of the other. But its conclusions were similar, so that the complaints the referee had levelled at the author's paper were applicable to his own. The charitable explanation is that the receipt of the original manuscript had prompted him to write up some work lying in his drawer, and that he had been pleased with the result. He appeared to be entirely innocent of the notion that he had allowed a conflict of

interest to get the better of him. For what it is worth, the original manuscript was eventually published, and the referee's returned.

That referees are regularly exposed to conflicts of this kind is inevitable. Are they not, after all, chosen for their reputations in their fields? But there is a good case to be made for believing that they should be readier to refuse to referee manuscripts that come too close to their own fields of interest.

That is certainly the case with an incident of a few months ago. A manuscript submitted for publication was sent to the head of a laboratory working mostly in the same field, who replied to say that the work was useful, but that it would be best placed in a more specialized journal. The case for saying that was not entirely self-evident, but the author was asked to comment. Before he had had a chance to do so, he saw in another journal a paper by a group from the referee's laboratory. By that time, the author had also recognized the referee's distinctive style.

"Why", was the complaint, "did not the referee disclose his interest, and disqualify himself?" If the referee and his colleagues had underestimated the interest of their original manuscript, and had then decided to send it to a specialized journal, would not the referee afterwards be prejudiced against publication in *Nature* of other papers apparently saying little more? Mercifully for the author's peace of mind, his paper was a little more substantial, and there was also further data to be added in response to technical comments from other referees. Eventually the text saw the light of day.

To be fair to the referee, what was asked of him must have created a dilemma. He could not properly have disclosed the existence of his colleagues' paper, or the imminence of its publication. That information (if relayed to the author) would have excited the complaint that the author was being fobbed off on the strength of a paper nobody else had seen. But he may have shrunk from using the phrase "conflict of interest" for fear of suggesting sinister happenings. But he should not have been so shy.

That is certainly the case of the person, a few months ago, who visited a colleague's laboratory only to find that photocopies of a paper he had submitted to *Nature* were in the hands of several people in the laboratory, who quizzed him carefully about the details. Apparently the author answered the questions without rancour, but afterwards he complained. Rightly, it will be thought. It is more than a little worrying that the photocopying referee, one of three referees, had

not submitted his report five weeks after receiving the manuscript, when, out of fairness to the author, it seemed proper to decide the fate of the manuscript on the strength of the two reports to hand.

Injured authors in such circumstances should not allow themselves to be misled by what seems to be the mystique of the publication process. That does not require that natural emotions should be suppressed. To have shown anger, even to have vented it, would not have been inappropriate.

A third recent lapse centres on another failure to declare a conflict of interest. An author's manuscript was rejected on the basis of reports from three referees, including one working in very much the same field whose report, while positive, was at best lukewarm. The manuscript was then submitted to another journal, whereupon it emerged that a similar paper had already been received and accepted for publication. At that stage, inevitably, the author guessed the identity of the referee.

Worse, the author raised the question whether the referee might have been helped by an early sight of his own manuscript. Technically, that might just have been possible. It might, of course, be argued that by that stage the problem was not *Nature*'s. One manuscript had been returned and the other never seen. But in the interests of general seemliness, it seemed proper to look into the issue. Mercifully, there was published evidence of the referee's earlier interest in the same problem. Moreover, the referee had a convincing tale to tell. But the clinching evidence was that he had returned his report within two days. It seems eventually to have been accepted by all concerned that nothing sinister occurred.

The moral in these tales is that there is no shame in a simple declaration of conflict of interest followed by a statement that the manuscript is being returned unread. Referees in such a case can be sure that they will not be asked to explain themselves — and if asked can say no.

But how can an author be sure that information in his manuscript never leaks out? There can be no certainty. The only assurance is that referees provide endless examples of studiously correct behaviour. But it is, for example, unthinkable that, if the Watson and Crick paper on DNA (*Nature* 171, 738; 1953) had been sent to referees, its substance would not have been going the rounds within days. In that case, it would not have mattered. In others, leakage can be a serious business.

John Maddox

Californians feel the tension

Tom Henyey

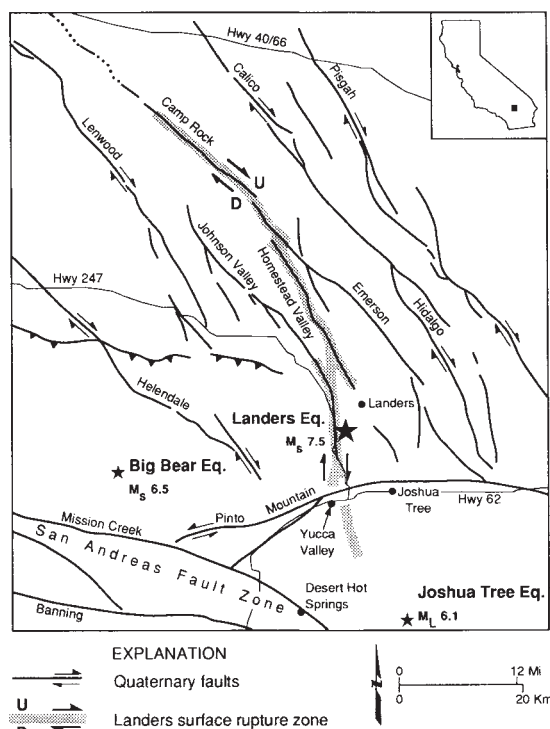
ON 28 June 1992 residents of southern California were awakened by the largest earthquake to strike the state in 40 years. Although unusually large, this quake was not on the well-known San Andreas fault (see figure) but rather it was near the town of Landers along a series of branching structures a short distance to the north. A number of questions immediately came to mind. Was this the 'big one'? Did the earthquake relieve any of the pent up stress on the San Andreas system? What, if any, might be the effect of this earthquake and its aftershocks on the San Andreas proper? Harris and Simpson address some of these questions on page 251 of this issue¹.

Specifically, they examine changes in the three-dimensional static crustal stress field attributable to the Landers earthquake, and the impact of these changes on neighbouring faults such as the San Andreas. Of greatest interest is the fact that although the Landers earthquake relieved crustal stress along the Landers rupture, this was not true for different segments of the San Andreas and other nearby faults. Some fault segments have actually been brought closer to failure by the Landers event, and as such there has been an increase in seismic hazard.

Modelling of static-stress changes accompanying earthquakes in an elastic medium is not new, although the application to changes in shear stress on neighbouring faults (Coulomb failure stress) is more recent^{2,3}. In the past, seismologists had not been particularly impressed by the levels of static-stress change following earthquakes, which are relatively small even at short distances from the rupture. Furthermore, the notion that the Earth can be treated as an elastic half-space or an elastic layer overlying a ductile lower crust as required by the models, had its sceptics, even though far-field strain measurements such as those being made at Pinon Flat in southern California (ref. 4 and D. Agnew, personal communication) would suggest otherwise — at least over human time-scales. But on contemplation, elasticity notwithstanding, a small increment of shear stress applied to a fault on the brink of failure might be the proverbial straw that breaks the camel's back.

Given the ripeness for failure of much of the southern San Andreas fault⁵, calculation of static stress field changes following the Landers earthquake was quickly recognized by seismologists as extremely relevant to evaluating the earthquake hazard in southern California. Stein, King and Lin, and Jaumé and

Sykes have made similar calculations to Harris and Simpson's, in papers appearing in *Science*^{6,7}. Each paper uses a somewhat different analytical approach to calculation of the Coulomb failure stress on neighbouring faults (an increase in the Coulomb failure stress brings a fault closer to failure, a decrease makes it safer), so that the final numbers



Quaternary faults and preliminary estimate of surface rupture associated with the Landers earthquake, San Bernardino County, California. Arrows indicate relative sense of displacement: U, up; D, down. (Figure courtesy of Geomatrix Consultants, San Francisco, California.)

are somewhat model dependent. Nevertheless, all three groups arrive at roughly the same conclusions. They are that the Coulomb failure stress was increased on the San Bernardino Mountains and Coachella Valley segments of the San Andreas fault as a result of the Landers earthquake, and decreased on the Mojave segment. Palaeoseismic data from the San Bernardino Mountains and Coachella Valley indicate that these segments are highly ripe for an earthquake of magnitude 7.5 or more. Large metropolitan areas including San Bernardino, Riverside, Palm Springs and Indio could suffer greatly.

In addition, both Harris and Simpson and Stein *et al.* show that the Coulomb failure stress was favourable for rupture on the causal fault of the magnitude 6.5 Big Bear earthquake, the principal aftershock to Landers which occurred about

3½ hours later. Stein *et al.* further show that a series of moderate-sized nearby earthquakes between 1975 and 1992 combined to produce a Coulomb failure stress that was favourable for the Landers rupture. Finally, all three groups recognize the potential for longer-term changes in the stresses due to both relaxation and pore-fluid processes. For example, Jaumé and Sykes suggest that pore-fluid pressure restoration following an earthquake may further increase the Coulomb failure stress with time.

A particularly intriguing question is whether these observations imply a change in seismic hazard and, if so, whether it can be quantified. To address this question, Harris and Simpson and Stein *et al.* introduce the notion of 'clock advance'. In effect, an increase in the horizontal shear stress that is beyond the underlying loading rate for a given fault translates into a clock advance, a decrease in the time to the next earthquake on that fault. Harris and Simpson compare the change in shear stress or Coulomb failure stress to the yearly stress load estimated using a simple model of long-term horizontal slip rate; Stein *et al.* estimate the change in occurrence time of the next major earthquake by comparing the immediate slip required to relieve the incremental shear to the annual slip rate. The results suggest that the clock has been advanced by 8–22 years and 2–14 years for the San Bernardino Mountains and Coachella Valley segments of the San Andreas fault, respectively. It has also been advanced for

the northern San Jacinto and Imperial faults, but has been delayed by 2–10 years for the Mojave segment of the San Andreas.

The final question is: does this change the reported probabilities of future large earthquakes on the San Andreas and San Jacinto faults? The answer depends on whether or not one believes that the probabilities previously determined⁵ depend on the occurrence of Landers-type events. For example, if Landers is inde-

1. Harris, R. A. & Simpson, R. W. *Nature* **360**, 251–254 (1992).
2. Stein, R. S. & Lisowski, M. J. *Geophys. Res.* **88**, 6477–6490 (1983).
3. Reasenber, P. A. & Simpson, R. W. *Science* **255**, 1687–1690 (1992).
4. Wyatt, F. K. J. *Geophys. Res.* **93**, 7923–7942 (1988).
5. Working Group on California Earthquake Probabilities *US Geol. Survey Open-File Rep.* 88–398 (1988).
6. Stein, R. S., King, G. C. P. & Lin, J. *Science* **258**, 1328–1331 (1992).
7. Jaumé, S. C. & Sykes, L. R. *Science* **258**, 1325–1328 (1992).

pendent of great San Andreas earthquakes, then the 1988 probabilities can be revised by defining an effective elapsed time since the last earthquake as the time it would take to accumulate aseismically the change in Coulomb failure stress on the San Andreas caused by the Landers earthquake plus the actual elapsed time. Doing this for the clock advances being reported now makes imperceptible changes in the one-year probabilities as reported in 1988. But it is important to note that the small

change in probability predicted by the clock advance method may not adequately reflect the change in seismic hazard. It is based on a simple quasiperiodic model for earthquake occurrence that may not capture the true physics of stress interactions following a large earthquake. □

Tom Henyey is at the Southern California Earthquake Center, University of Southern California, Los Angeles, California 90089, USA.

EPIDEMIOLOGY

Molecular potential

Colin Garner

MOST human carcinogens have been identified by epidemiological studies or by astute clinicians who have noticed clusters of cancer cases resulting from particular exposures¹. Classical epidemiological studies suffer from a number of disadvantages, however, including poor sensitivity and lack of resolving power. The paper by Perera *et al.*, on page 256 of this issue², represents a new approach to establishing carcinogen exposure, where population studies are conducted at a molecular level. It stems from the general hypothesis that organic chemical carcinogens, such as benzo(a)pyrene and other organic pollutants, are activated in the body to chemical species that interact with biological molecules. Such interactions give rise to DNA-bound moieties (adducts), chromosome aberrations, sister chromatid exchanges and oncogene activation³.

Perera *et al.* looked at six different endpoints of genotoxic damage in blood obtained from two Polish populations — urban, where air pollution was high, and rural, where pollution was some ten times lower. Summer–winter comparisons were also made in the same individuals (surely the best experiments to conduct, because individuals act as their own controls). For the majority of parameters measured there were statistically significant differences between winter and summer exposures in the urban and rural groups, most notably so for polycyclic aromatic hydrocarbon adducts measured in blood lymphocytes.

The authors did not perform a complete summer–winter–summer cycle in either population, which might have shown an increase in the level of biomarker studied in the winter and a subsequent fall the following summer. Moreover, for some endpoints such as chromosome aberrations, too few blood samples were analysed to arrive at definite conclusions. But what can be stated

is that there were highly significant differences in biomarker amount, with winter exceeding summer and urban exceeding rural, which correlated with the rise in atmospheric air pollution.

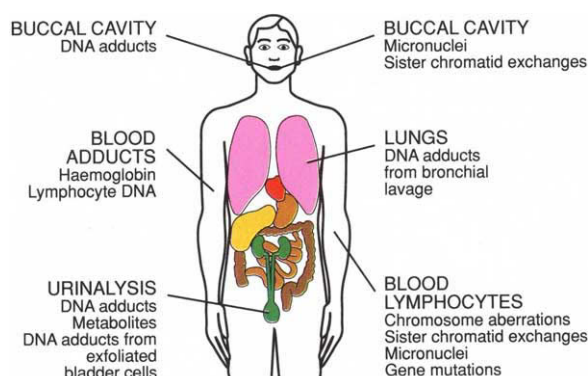
If molecular epidemiology is to be of practical use, it should be able to identify new cancer risks long before cancer develops; it should be cheaper than classical epidemiological studies (otherwise why bother?); and it should assist in ranking cancer hazards in terms of potency. Molecular epidemiology should not be confused with the measurement of internal dose (the dose of pollutant, for example, that an individual absorbs). Here the measurement made may have absolutely nothing to do with the eventual biological outcome — cancer.

Again, molecular epidemiology studies should ideally be predictive of future cancers so that, as Perera *et al.* correctly state, intervention studies can be performed in order to examine the reduction in cancer risk as measured using the appropriate biomarker. Many of the molecular epidemiological endpoints currently being examined reflect only recent, not cumulative, exposure to carcinogens. The figure shows some of the endpoints currently being used to study carcinogen exposure at a molecular level: each has its strengths and weaknesses. For instance, on their own, chromosome aberrations are poor indicators of a particular carcinogen exposure because we do not know what agent(s) were responsible for the aberration observed.

In an elegant series of experiments, Bates and his colleagues⁴ examined a group of workers exposed to the geno-

toxin, ethylene oxide. This chemical provides a good model for study because it does not require metabolic activation, hence individual differences in cytochrome P-450 are likely to be unimportant, and there are well-established physico-chemical methods for the characterization of protein adducts. Bates *et al.* found that the sensitivity of the various biological endpoints in blood obtained from exposed workers was, in ascending order of sensitivity, lymphocyte gene mutations, micronuclei, chromosome aberrations, sister chromatid exchanges, haemoglobin adducts. Unfortunately, DNA adducts were not measured, but these are likely to be the most sensitive biomarker and perhaps the most relevant.

If the field of molecular epidemiology is to advance, certain fundamental questions will have to be answered. For genotoxic compounds, that is, chemicals such as those found in polluted atmospheres, we need to know several things (for review and comment see ref. 5). First, what the relationship is between protein adducts measured in the blood and DNA adducts in a target organ, such as the lung; second, what the relationship is between DNA adducts and cancer; third, whether animal models can be used to make predictions about



Non-invasive targets for human biomonitoring.

humans; and finally, whether measurements of genotoxic damage will be helpful in intervention studies. There is a tendency to measure any biomarker and try and correlate it with another biomarker, or with some disease outcome, without full calibration of the system or establishing its relevance.

What is needed is a more systematic approach which takes advantage of known carcinogen exposures, such as cigarette smoking, and relates them to appropriate biomarkers⁶. Many laboratories have examined the DNA adduct levels in various organs throughout the body in cigarette smokers⁶. Although there is a correlation between levels of DNA adducts in the lung and the numbers of cigarettes smoked⁷, it is much harder to demonstrate a similar correla-

Lord Ashby (1904 – 1992)

With the death of Eric Ashby on 22 October, the university world has lost its most eminent statesman. Respected as much in the United States and the countries of the British Commonwealth as he was in Britain, Ashby had a profound influence on the character of higher education in the crucial years from the 1950s to the early 1970s.

Eric Ashby was born in Bromley, Kent, and educated at the City of London School and Imperial College London, where he graduated in botany. After various more junior appointments, in 1938, at the age of 34, he gained the chair of botany at the University of Sydney in Australia. He made an immediate mark upon his department and his remarkable skills in administration were soon recognized. With the advent of war, he took on wider national responsibilities, serving as chairman of the Australian National Research Council, special adviser to the Australian prime minister on the mobilization of scientific resources for war, director of the Scientific Liaison Bureau and subsequently as counsellor and chargé d'affaires at the Australian Legation in Moscow, an episode described in his book *Scientist in Russia* (1947).

Ashby returned to England in 1946 as Harrison Professor of Botany at Manchester, and his leadership contributed to the rapid rise of the department there to become one of the foremost in the country. Four years later, he was appointed president and vice chancellor of Queen's University, Belfast, where his energy, vision and persuasiveness transformed the institution. He recruited a new generation of outstanding young professors, built a remarkable sense of community and gave the growing university both the facilities and the support that it needed. He later served as its chancellor from 1970 to 1983.

Ashby's next appointment, to the mastership of Clare College, Cambridge, caused some to wonder whether a man who had been to neither Oxford nor Cambridge could provide the same successful leadership at Clare that he had at Belfast. He quickly showed that he could: Clare's student body became



more representative — it was one of the first Cambridge colleges to admit women — and it excelled in its degree results. It was at his patient urging that the college decided to devote a substantial part of its endowment to establishing a new college, Clare Hall, for graduate students and visiting scholars. Ashby served as vice chancellor of the University of Cambridge from 1967 to 1969, steering the university through those turbulent times with moderation, courage and civility.

Ashby's contribution to the world of learning increased in scope with time: there was scarcely any influential national committee on higher education or science of which he was not a member or chairman, ranging from the Advisory Council on Scientific and Industrial Research to the presidency of the British

Association for the Advancement of Science. Nor was his interest limited to Britain: he was head of the commission that redesigned post-secondary and higher education in Nigeria and he was the only non-US member of the influential Carnegie Commission on Higher Education.

In his later years, he returned to his earlier ecological interests and employed them in his work as a member of the Royal Commission on Environmental Pollution and as chairman of the working party on pollution control in connection with the UN Stockholm Conference. His two books on environmental issues, *Reconciling Man with the Environment* (1978) and *The Politics of Clean Air* (with Mary Anderson, 1981) are models of balance and responsibility.

Behind Lord Ashby, the quintessential public servant, the perfect committee chairman, was Eric Ashby, the private man who commanded the highest affection. He lived modestly, loving his books, his writing and especially his chamber music. He was a skilled viola player and when, in his mid-eighties, arthritis prevented him from playing the viola to his own high standard, he enthusiastically took up the electronic keyboard as a substitute.

In his *Portrait of Haldane* (written with Mary Anderson in 1974), Ashby wrote: "... much of Haldane's achievement in public life ... was not due to the direct exercise of power on his part, but to the subtle way in which he could infect other people with his enthusiasm, persuade them to his views, equip them with his advice, and enlist their loyalty and affection". It is a description that serves as a fitting epitaph to Eric Ashby.

Frank H. T. Rhodes

Frank H. T. Rhodes is president of Cornell University, Ithaca, New York 14853, USA.

tion for organs that are less at risk of developing cancer. Most frustratingly, if individual adduct levels are high in the lung they need not be so in another organ, such as the bladder⁸. It is hardly surprising, therefore, that it has been difficult to correlate white blood cell adduct levels in smokers and adduct levels in the lung of the same individual⁹.

There have been successes, however. Molecular epidemiology has been able to demonstrate that carcinogen exposure in humans gives rise to the same type of genotoxic damage that occurs in carcinogen-exposed experimental animals. So we must assume that animal models have some relevance to human cancer.

What then is the way forward? We should be selecting populations that are particularly at risk of developing cancer,

such as cigarette smokers or certain patients undergoing chemotherapy, and should be looking for biomarkers that are predictive of the disease. This can best be achieved for easily obtainable cells, for example, exfoliated bladder cells, which are voided in the urine, skin cells or cells from the mouth. Aetiologic agents for these cancers are cigarette smoking for the bladder and mouth, and ultraviolet light for the skin. Establishing appropriate biomarker levels in these cells and correlating these with some early event in the cancer process, such as a mutation in *ras* or another oncogene, may provide the best means of determining if molecular epidemiology will replace conventional epidemiology as a way of identifying and controlling cancer risks. Only when events in the target organ have been unravelled can we search for surrogate biomarkers.

The significance of the paper of Perera *et al.* is that it points to what may be possible. It will not help in cancer risk assessment, but rather in internal (as opposed to target dose) measurement, and it will be a valuable study for those working on pollution control. □

Colin Garner is in The Jack Birch Unit for Environmental Carcinogenesis, Department of Biology, University of York, Heslington, York YO1 5DD, UK.

1. Doll, R. *Nature* **265**, 589–596 (1977).
2. Perera, F. P. *et al.* *Nature* **360**, 256–258 (1992).
3. Harris, C. C. *Cancer Res. Suppl.* **51**, 5023s–5044s (1991).
4. Tate, A. D. *et al.* *Mutation Res.* **250**, 483–497 (1991).
5. Garner, R. C. *et al.* (eds) *Human Carcinogen Exposure: Biomonitoring & Risk Assessment* (IRL, 1991).
6. Beach, A. C. & Gupta, R. C. *Carcinogenesis* **13**, 1053–1074 (1992).
7. Phillips, D. H. *et al.* *Nature* **336**, 790–792 (1988).
8. Routledge, M. N. *et al.* *Mutation Res.* **282**, 139 (1992).
9. Van Schooten, F. J. *et al.* *Carcinogenesis* **613**, 987–993 (1992).

A looming Arctic ozone hole?

J. D. Mahlman

INCREASING concentrations of greenhouse carbon dioxide in the atmosphere may expose us to the possibility of an Arctic version of the Antarctic 'ozone hole', a hazard we have avoided so far because of the special atmospheric conditions of the Northern Hemisphere. But calculations by Austin, Butchart and Shine, on page 221 of this issue¹, show that the changes to be expected from the almost inevitable doubling of CO₂ levels² may undo that protection.

The 1985 discovery of a very large springtime loss of lower-stratospheric ozone over Antarctica was a turning point in atmospheric chemistry. Considering that this ozone hole was totally unpredicted, it is impressive that its essence was explained so quickly. Successful merging of ground-based and satellite measurements, polar aircraft expeditions and atmospheric modelling was crucial to our current understanding.

This understanding is centred about a remarkable set of chemical processes initiated by the formation of polar stratospheric clouds (PSCs), composed of water ice and nitric acid trihydrate, in the deep of polar winter³. The main chemical cause of ozone destruction results from catalytic cycles involving chlorine and bromine liberated from chemically inert reservoir molecules by heterogeneous reactions on the PSC surfaces. But PSC formation depends very sensitively on the temperature in the polar night, itself controlled, in essence, by infrared cooling and by stratospheric dynamical activity (essentially the upward propagation of meteorological disturbances from the troposphere). With higher disturbance levels, the polar temperatures become warmer. Ozone losses are much smaller over the Arctic than over the Antarctic because the Northern Hemisphere is dynamically much more active than the Southern Hemisphere.

But will this always be so? Austin *et al.* note that the stratosphere is expected to become cooler owing both to CO₂ increases and to long-term ozone decreases^{4,5}. This should increase the likelihood of PSCs forming over the Arctic because the saturation vapour pressure of gas-phase water and nitric acid are strong exponential functions of temperature. With extensive PSCs, increased chlorine- and bromine-driven ozone destruction can be expected over the Arctic region.

The threat of an ozone hole over the heavily populated Northern Hemisphere will undoubtedly stimulate a lot of concern, probably more than the already

present Antarctic hole. How reliable is the prediction? Austin *et al.* have almost certainly done the best that can be achieved at present. For more reliable calculations we will probably need at least five years' further intensive research into the pertinent atmospheric chemistry and dynamics.

The stratospheric cooling effect of added CO₂ described by Austin *et al.* is beyond doubt. It depends only on the straightforward physics of infrared radiative transfer. Also the ozone reductions that have already occurred in the lower stratosphere have the effect of introducing their own amplifying feedback⁶. That is, ozone reduction induces cooling (less solar ultraviolet is absorbed), which produces more PSCs, liberating more chlorine and bromine... Austin *et al.* suspect that this is potentially the predominant process. How it will play out in the real world awaits improved modelling and a better understanding of the governing physical processes.

Additional complications arise in the details of ozone chemical processes acting at and near the edge of the stratospheric vortex that forms in the Arctic winter. The net efficiency of Arctic ozone-destroying chemistry could be substantially influenced by the process of shredding and tearing off strips and chunks of chemically processed air by disturbances impinging on the vortex edge⁷. But with modelling of these processes still in its infancy, the quantitative effects are currently beyond calculation.

Long-term increases in the water-vapour content of the stratosphere could make an Arctic ozone hole more likely. The stratosphere's internal source of water vapour is oxidation of methane, the concentration of which is currently increasing by a little less than 1 per cent a year². Increasing amounts of water vapour mean a higher condensation temperature for PSC formation. But water is also imported into the stratosphere from the troposphere in equatorial regions, passing through the cold tropopause on the way. Climate models with increasing greenhouse gases predict a warming of that region², so that less water vapour should be trapped out by the cold as tropospheric air passes through. On the other hand, observations suggest that the tropical tropopause has actually cooled over the past two decades², possibly because of polar ozone reduction penetrating to the lower-latitude tropopause region⁸.

A subtle, but potentially significant factor not considered by Austin *et al.* is the effect tropospheric climate change

might have on the dynamical forcing of the stratosphere. Currently, there are two perspectives on this. My view is that a greenhouse-warmed troposphere would probably reduce the amplitude of tropospheric weather disturbances. This is because the simulated reduced meridional temperature contrast near the Earth's surface and increased tropospheric water vapour content in our doubled-CO₂ calculations together reduce the amplitude of tropospheric planetary waves. In addition, increased CO₂ amounts increase the rate at which radiative transfer brings perturbed stratospheric temperature back to radiative balance. Both of these effects would increase the cooling of the Arctic polar vortex.

This view contrasts with that presented by a Goddard Institute of Space Studies (GISS) model study, which predicts greater dynamical forcing of the stratosphere in a doubled-CO₂ climate⁵. Suffice it to say that the discrepancy appears to arise from differences in the ways that the calculations treat unresolved convection and gravity-wave processes. If our calculations are correct, reduced dynamical activity would increase the likelihood of an Arctic ozone hole. If the GISS calculations are correct, a process to stave it off may have been identified. Increased dynamical forcing for the winter stratosphere could keep the polar regions warm enough to reduce PSC-induced ozone destruction.

As in many studies of climate change, detection of a trend towards an Arctic ozone hole will not be simple. The Arctic winter varies greatly from year to year. Even with stratospheric cooling, occasional warmer polar winters will buck the trend. Nevertheless, the chances of large ozone-destroying events, like those already known from Antarctica, could be growing.

When might we see such an Arctic ozone hole? As significant Arctic ozone losses have already been identified⁹, one could argue that the process is already well underway. Even with the Montreal Protocol, stratospheric chlorine and bromine loading may exceed 1980 values (which were sufficient to trigger the

1. Austin, J., Butchart, N. & Shine, K. P. *Nature* **360**, 221–225 (1992).
2. Houghton, J. T., Callander, B. A. & Varney, S. K. (eds) *Climate Change 1992, The Supplementary Report to the IPCC Scientific Assessment* (Cambridge University Press, 1992).
3. Solomon, S. *Nature* **347**, 347–354 (1990).
4. Fels, S. B., Mahlman, J. D., Schwarzkopf, M. D. & Sinclair, R. W. *J. Atmos. Sci.* **37**, 2265–2297 (1980).
5. Rind, D., Suozzo, R., Balachandren, N. K. & Prather, M. J. *J. Atmos. Sci.* **47**, 475–494 (1990).
6. Ramaswamy, V., Schwarzkopf, M. D. & Shine, K. P. *Nature* **355**, 810–812 (1992).
7. Schoeberl, M. R., Lait, L. R., Newman, P. A. & Rosenfield, J. E. *J. geophys. Res.* **97**, 7859 (1992).
8. Mahlman, J. D., Pinto, J. P. & Umscheid, L. J. *J. Atmos. Sci.* (submitted).
9. WMO/UNEP *Global Ozone Research and Monitoring Project, Report No. 25* (World Meteorological Organization, Geneva, 1992).

Antarctic ozone hole) well into the middle of the twenty-first century⁹. Projections of CO₂ levels suggest a doubling of preindustrial values after the middle of the next century². Beyond that period, chlorine and bromine concentrations will slowly decrease (over a 100-year timescale), while CO₂ levels will probably continue to increase². It remains difficult to pinpoint onset or disappearance

dates. Nevertheless, it is safe to assume that the threat of an Arctic ozone hole will be with us for a long time unless CO₂, chlorine and bromine emissions are brought substantially below currently projected levels. □

J. D. Mahlman is in the Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, New Jersey 08542, USA.

TUMORIGENESIS

Suppression with a difference

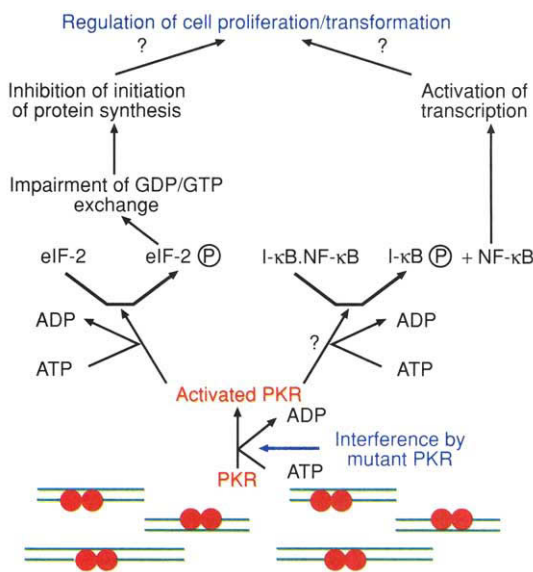
Mike Clemens

THE concept of tumour suppressor genes (sometimes called anti-oncogenes) is now a familiar one. Mutation or loss of both copies of sequences such as the retinoblastoma (*RB*) or *p53* genes is associated with a number of human cancers¹, suggesting that their protein products keep in check an otherwise uncontrollable ability of cells to proliferate. Although there are several other DNA sequences that are candidates for such genes², we still know of far fewer tumour suppressor genes than oncogenes and much less about how their products work. Now, however, there is evidence from two groups^{3,4} that a well-characterized protein kinase is a tumour suppressor.

The enzyme in question is the double-stranded RNA (dsRNA)-regulated protein kinase variously known as p68, DAI (for dsRNA-activated inhibitor), dsI, P1 kinase and PK_{ds}. In accord with an informal agreement made by many workers in the field, I will here call it PKR (for RNA-activated protein kinase). The enzyme is inducible by interferons, and is also present at a low level constitutively in many cell types. It has a major role in one of the pathways by which the antiviral effects of the interferons are exerted in infected cells⁵ (see figure). The idea is that dsRNA produced during the replicative cycles of a range of viruses activates PKR, which then goes on to inhibit protein synthesis (both viral and cellular) by phosphorylating a subunit of the essential polypeptide chain initiation factor eIF-2. As a result, the exchange of GDP for GTP on eIF-2 (catalysed by another factor) is blocked, chain initiation is impaired and virus replication is thereby inhibited.

So what has this got to do with tumour suppression? The complementary DNA for human PKR was cloned two years ago⁶ and several laboratories have since been busily making site-directed mutants and expressing these proteins in a variety of systems. What the new reports show is that 3T3 fibroblasts stably over-

expressing forms of inactive PKR are highly tumorigenic when injected into nude mice. The protein kinase was inactivated either by mutation of a single residue involved in the phosphate transfer reaction⁴ or by deleting a short



Possible mechanisms of action of PKR in the control of cell proliferation and transformation. PKR monomers (red) are activated by autophosphorylation following their binding to dsRNA (green; protein dimerization may be required). Active PKR phosphorylates the protein synthesis initiation factor eIF-2 and inhibits translation; it is also possible that PKR phosphorylates I-κB, the inhibitor of the transcription factor NF-κB, causing dissociation of I-κB from NF-κB and activation of the transcription of a number of genes stimulated by NF-κB. Interference with the activation or activity of PKR by mutant forms of the protein kinase may interfere with one or both pathways, leading to tumorigenesis by steps that remain to be determined.

stretch of amino acids probably required for substrate recognition³. Neither cells transfected with a plasmid encoding the wild-type enzyme nor cells containing the plasmid vector alone were tumorigenic. So it seems likely that the overexpression of inactive PKR interferes with the functioning of wild-type PKR in a dominant negative fashion leading to tumour development.

The results are exciting for several reasons. First, they identify a potential tumour suppressor activity that is distinct from those currently known. Both the *RB* and *p53* gene products are believed to function as DNA-binding proteins and/or components of transcription complexes, directly regulating the expression of downstream genes. By contrast, PKR is a cytoplasmic, ribosome-associated protein that apparently acts by regulating translation. Second, the evidence for the involvement of PKR in tumorigenesis may at least partly explain the effects of the interferons as cell-growth inhibitors and anti-tumour agents⁷.

Third, we know of a number of viruses that have evolved mechanisms for the inactivation or degradation of endogenous PKR⁸. Such mechanisms include production of small RNA molecules that prevent the activation of the enzyme by dsRNA^{9,10}, synthesis of proteins that sequester the dsRNA activators¹¹, and

even the production of a decoy protein that looks rather like the enzyme's substrate, the α-subunit of eIF-2 (ref. 12). In the case of viruses that do not kill their host cells but instead produce a mitogenic response or transform the cells to a tumorigenic phenotype, part of the growth-promoting effect might involve the inactivation of PKR by viral gene products. Such an effect, if it occurs in immortalized cells, would be analogous to that of the viral proteins E1a (adenovirus), T antigen (simian virus 40) and E7 (human papillomavirus), which bind to and interfere with the action of the *RB* tumour suppressor protein¹.

How might a high level of mutant PKR interfere with the activity of the endogenous protein kinase? One simple explanation would be that the two proteins compete for binding of limiting amounts of dsRNA activators (the nature of which is largely unknown). Alternatively, given that there is some evidence that PKR exists as a homodimer and that this configuration may be necessary for its activation by autophosphorylation in the presence of

dsRNA, the formation of heterodimers between active and inactive enzyme subunits could interfere with the activation process.

Whatever the mode of action of the PKR mutants, it might seem surprising that the elimination of a negative translational control mechanism can lead directly to a tumorigenic phenotype. There is a precedent, however, in that

over-expression of a protein synthesis factor, the cap-binding protein eIF-4E (4F α), which promotes translation, also transforms cells^{13,14}. A possible mechanism in either case could involve enhanced rates of synthesis of proteins critical for cell-cycle control. Nevertheless, the ability of cells expressing mutant PKR to phosphorylate their eIF-2 has been reported not to be impaired¹⁵.

There is a possibility that the mutant PKR affects another pathway altogether, regulation of the NF- κ B family of transcription factors¹⁶. Results from Bryan Williams's laboratory (Cleveland, Ohio), presented at the meeting of the International Society for Interferon Research in Toronto in September, suggest that PKR can phosphorylate at least one member (MAD-3) of the I- κ B family of inhibitors that control NF- κ B activity, leading to the dissociation of this inhibitory subunit from the NF- κ B complex and activation of the latter for DNA binding. NF- κ B can be activated by dsRNA in intact cells¹⁷; if these effects of PKR are applicable *in vivo*, one can then envisage a model in which impaired activity of NF- κ B in cells over-expressing mutant PKR results in an inappropriate pattern of gene expression and in loss of growth regulation. There are precedents for disruptions in the I- κ B/NF- κ B system leading to tumorigenesis, in that the *rel* oncogene is a member of the NF- κ B family¹⁶.

We can now look forward to intense activity in the study of PKR, on both the translational and transcriptional fronts. Whatever the final outcome, it seems that the versatility of a protein kinase has once again led us in a totally unexpected direction. □

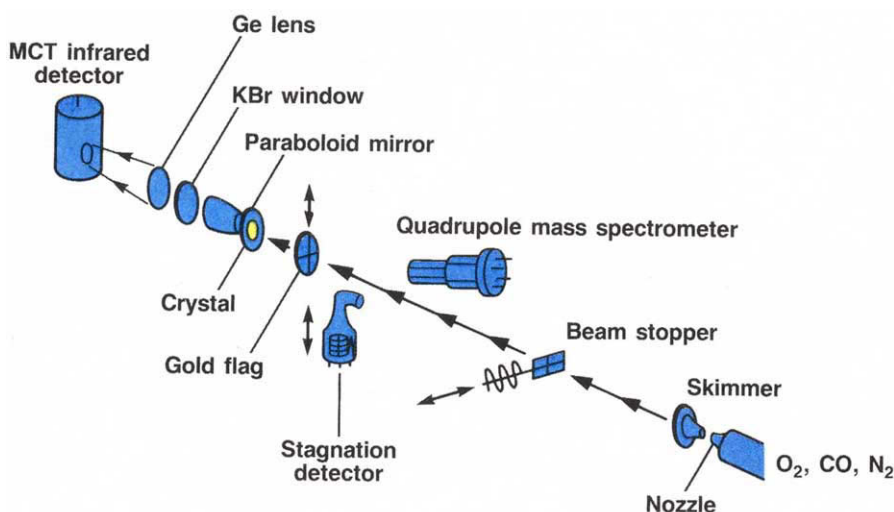
Mike Clemens is in the Division of Biochemistry, Department of Cellular and Molecular Sciences, St George's Hospital Medical School, London SW17 0RE, UK.

A little heat goes a long way

John T. Yates Jr

A REMARKABLY sensitive method for detecting the heat evolved when a gas is adsorbed onto a crystal surface is described by D. King and coworkers on page 243 of this issue¹. They use their method to find out why alkali metals are effective at promoting the activity of certain industrial catalysts. It turns out that alkali promotion greatly increases the heat evolved when CO is adsorbed onto nickel; this increase can be related

King's group vaults this barrier to heat-of-adsorption measurements using a pulsed molecular beam of adsorbate, incident on a single crystal film only 2,000 Å thick³. A mercury-cadmium-telluride detector measures the crystal's temperature rise during a sequence of gas pulses, each containing only 10¹² molecules, by measuring the faint infrared glow developed by the warmed crystal (see figure). These measurements



The new calorimetric apparatus developed by King and coworkers. The high-velocity beam is chopped (stopper) to give a pulsed source of gas, incident on a 2,000-Å-thick single crystal. Infrared radiation from the warmed crystal is collected by special optics and measured by a mercury-cadmium-telluride (MCT) detector. The stagnation detector, gold flag and quadrupole mass spectrometer are used in measuring the efficiency of adsorption on the crystal.

to the way the gas molecules bind to the surface.

Despite the vast arsenal of specialized tools available to surface scientists, an accurate calorimetric method has been lacking. In the early days of surface chemistry, heats of adsorption were typically measured by calorimetry². But because it works only when measuring rather large amounts of heat, the calorimetric method could not be accurately used for the study of well-defined single crystal surfaces, which generally have a surface area of only 1 cm², containing about 10¹⁵ adsorption sites. Instead, powders with high-surface area but ill-defined surface structure or thin films were employed as adsorbents for calorimetry. Thus, calorimetry was blind to the variations in surface bonding which occur on different single crystal planes for a given solid material. Other methods such as the measurement of the 'isosteric' heat of adsorption were adopted. This method requires that the adsorbate and gas be at equilibrium, which is not necessary for the calorimetric method of King.

now permit detailed insight into adsorption energetics on well-defined single crystal surfaces. Such work on metal surfaces provides important basic concepts useful to fields such as heterogeneous catalysis and corrosion passivation, and potentially in semiconductor surface chemistry.

King and coworkers have now applied their method to an important problem in the area of heterogeneous catalysis which has been a central focus of research in surface science. This has to do with the widespread use of adsorbed alkali metals to promote the catalytic activity for carbon monoxide chemistry on transition-metal catalysts. The central question is whether the alkali metals promote C-O bond breaking, and if so, at what temperature. Obtaining this understanding is important in understanding alkali-promoted Fischer-Tropsch catalytic chemistry, which converts CO and hydrogen to hydrocarbons. Ammonia synthesis also involves alkali promotion of metal catalysts.

King and coworkers find that the adsorption of CO on Ni(100) crystals

1. Marshall, C. J. *Cell* **64**, 313-326 (1991).
2. Sager, R. *Curr. Opin. Cell Biol.* **4**, 155-160 (1992).
3. Koromilas, A. E., Roy, S., Barber, G. N., Katze, M. G. & Sonenberg, N. *Science* **257**, 1685-1689 (1992).
4. Meurs, E. F. et al. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
5. Pestka, S., Langer, J. A., Zoon, K. C. & Samuel, C. E. *A. Rev. Biochem.* **56**, 727-777 (1987).
6. Meurs, E. F. et al. *Cell* **62**, 379-390 (1990).
7. Balkwill, F. R. *Cytokines in Cancer Therapy* (Oxford University Press, 1989).
8. Katze, M. G. *J. Interferon Res.* **12**, 241-248 (1992).
9. Mathews, M. B. & Shenk, T. J. *Virology* **65**, 5657 (1991).
10. Gunner, S., Rice, A. P., Robertson, H. D. & Mathews, M. B. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8687 (1990).
11. Imani, F. & Jacobs, B. L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7887-7891 (1988).
12. Davies, M. V., Elroy-Stein, O., Jagus, R., Moss, B. & Kaufman, R. J. *J. Virol.* **66**, 1943-1950 (1992).
13. Lazaris-Karatzas, A., Montine, K. S. & Sonenberg, N. *Nature* **345**, 544-547 (1990).
14. De Benedetti, A. & Rhoads, R. E. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8212-8216 (1990).
15. Meurs, E. F. et al. *J. Virol.* **66**, 5805-5814 (1992).
16. Nolan, G. P. & Baltimore, D. *Curr. Opin. Genet. Devel.* **2**, 211-220 (1992).
17. Visvanathan, K. V. & Goodbourn, S. *EMBO J.* **8**, 1129-1138 (1989).

with a partial monolayer coverage of potassium involves an abnormally high heat of adsorption. The magnitude of the increase in the heat of chemisorption, compared with the clean Ni(100) surface, cannot be explained solely by electronic effects in the CO molecule caused by electric fields near the polarized alkali atoms. There is an additional effect on the alkali atoms, in which electrical screening of the alkali atoms from each other by intervening CO molecules causes additional alkali polarization, evolving additional heat.

Thus, the high heat of adsorption of CO is the sum of an alkali-induced increase in the CO binding energy and a CO-induced increase in binding energy of the alkali atoms. Below 400 K, the dissociation of CO is not responsible for the high heat of CO adsorption, in agreement with our own work involving vibrational spectroscopy⁴ and laser-induced desorption of isotopically substituted CO molecules from nickel single-crystal surfaces containing potassium⁵. It is believed that islands schematically including $\dots K^+ CO^- K^+ \dots$ entities are involved at intermediate coverages of K and CO. At low coverages of K, we have observed a more local island interaction between K^+ and neighbouring CO molecules^{4,6}.

What do these results imply for the future? First of all, they clearly demonstrate the unique power of a sensitive new tool, which may be applied to a wide range of problems where adsorption energies need to be measured under controlled surface structural conditions. We should now be able to tie heat-of-adsorption measurements to other studies in surface science that involve atomic-level measurements on single crystals. Second, it might be possible to use King's method to measure heat effects between chemisorbed reactants as the temperature is raised enough to overcome activation barriers that inhibit surface reactions. If this could be done, calorimetry on single crystal films will offer much promise as a general surface-chemistry research technique, just as differential scanning calorimetry techniques have contributed much to our understanding of kinetics and thermochemistry in solid-state sciences. □

John T. Yates Jr is in the Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.

Following iceberg footprints

Eystein Jansen

LAYERS of sediments rich in glacial deposits, discovered a few years ago on the Atlantic sea bed, turn out to mark the trails of swarms of icebergs originating from ice sheets on North America in the depths of the last ice age, up to 40,000 years ago. Analyses by Bond *et al.*, on page 245 of this issue¹, show that these Heinrich deposits (named after their discoverer², who is also a co-author of the new work) extend 3,000 km across the North Atlantic, attesting to the remarkable range of the iceberg episodes.

Bond *et al.* trace the sources of the layers to limestone and dolomite rocks in northern Canada, and show that most of the layers are formed in a distinct band dispersed across the North Atlantic between Newfoundland and Europe at 40–50° N. Knowing the origin and dispersal pattern solves two important questions related to the Heinrich layers. First, each deposit must reflect a sudden surge of the margin of the Laurentide/Inuitan ice-sheet complex, which covered much of North America during the last ice age. Second, the deposition pattern is tied to the prevailing ocean current system at the time of formation. Melting occurred predominantly in the southern part of the glacial subpolar gyre in contact with warmer waters to the south. This is in line with the previous perception of the surface circulation pattern in the glacial North Atlantic³.

Now, knowing that sudden surges of the North American ice sheet took place, what was their fundamental cause? Bond *et al.* do not resolve this question, but point to some clues. There was a pronounced cooling of the surface waters when the Heinrich layers formed. The layers reflect brief but massive discharges occurring within periods of extreme cooling of the ambient North Atlantic Ocean. The cooling apparently had a strong impact on the ice sheets, and one may speculate that rapid ice-sheet growth may have made the ice-sheet margins more unstable and prone to calve and to produce sudden surges, releasing enormous numbers of icebergs into the ocean. The subsequent melting of the icebergs drastically diluted the surface waters of the North Atlantic with fresh water, documented by a strong reduction in the ¹⁸O content of carbonate fossils formed at the same time (glacial meltwater contains far less ¹⁸O than ocean water). Such dilution may provide a strong feedback on climate because the low salinity will stabilize surface waters and stop them sinking to form deep water. This, in turn, reduces ocean-atmosphere heat flux and

possibly alters the ocean-atmosphere CO₂ balance.

The massive changes in ocean surface properties indicate that the Heinrich events can be viewed as a potentially strong mechanism for rapid climate change. Certainly they fit into an emerging picture of climate instability during the last ice age. Sudden changes in North Atlantic temperatures have been identified from the deglaciation that led to the present warm climate^{4,5}. And isotope signals just reported⁶ reveal other sudden climate switches further back into the last ice age. These changes may have occurred over less than a human lifespan. But there appear to have been many more swings in North Atlantic climate than there are Heinrich layers. What were the deciding features that caused the layers to form in some cases but not in others? Until a detailed correlation with the new ice-core records from Greenland is performed, the impact of the Heinrich episodes on global and regional climate will remain unclear.

There is increasing evidence that the growth of ice sheets to their maximum extent was not gradual, but happened over just a few thousand years⁷. This evidence indicates that there has to be a dynamic ocean delivering moisture to high latitudes to promote drastic growth of the ice sheets, despite the ocean surface cooling which would tend to reduce moisture supply. Future research needs to find the exact phasing of ocean changes and corresponding ice-sheet responses. This will now be possible by tracing the Heinrich layers directly to their sources on the shelves. The accurate age determination now possible with accelerator dating and the phenomenal possibility of detailed core-to-core correlation via the Heinrich layers should enable researchers to tie land and ocean records of climate change together with unprecedented detail.

Bond *et al.* find that the content of calcareous fossils is low in the Heinrich layers, and suggest that the events lowered ocean biological productivity. In contrast, Sancetta, also in this issue⁸, concludes that times of greater iceberg activity in the glacial period were accompanied by high productivity. This is based on the ages and distributions of diatom assemblages, indicative of productive regions. The apparent contradiction is not insuperable. Many highly productive areas are characterized by diatom (silica) rather than carbonate sedimentation. This lends most credibility to estimates based on silica fossils. As the duration of the Heinrich events is

1. Al-Sarraf, N., Stuckless, J. T. & King, D. A. *Nature* **360**, 243–245 (1992).
2. Trapnell, B. M. W. *Chemisorption* 140–169 (Butterworths, London, 1955).
3. Borroni-Bird, C. E., Al-Sarraf, N., Andersson, S. & King, D. A. *Chem. Phys. Lett.* **183**, 516–520 (1991).
4. Ng, L. *et al.* *J. chem. Phys.* **86**, 6523–6530 (1987).
5. Sinniah, K., Sands, W. D., Yates, J. T. Jr & Janda, K. C. *J. Am. chem. Soc.* **113**, 3684–3688 (1991).
6. Alvey, M. D., Lanzillotto, A. M. & Yates, J. T. Jr *Surface Sci.* **177**, 278–290 (1986).

RÉSUMÉ

unknown, the low abundance of carbonate fossils may reflect only the dilution of the fossil flux by ice-rafted sediment particles. It may also be that the freshwater layer accompanying Heinrich events renders the environment more hostile to calcareous plankton than to diatoms.

Despite the wealth of new information on the Heinrich layers now becoming available, there are many unanswered questions. Indications that similar sediment layers are found further north and can be traced to sources in the Fennoscandian, Barents and Svalbard ice sheets need to be put on solid footing. If Heinrich layers were deposited from other ice sheets, at the same time as the ones in the North Atlantic, this will be an important new facet of ice-age dynamics in general.

The pacing of the Heinrich layers is different from the changes in the Earth's orbit that lie behind the Milankovitch mechanism of climate change, but Bond *et al.* are perhaps hasty in dismissing

orbital forcing as a cause of the events. The instability responsible evidently requires that ice sheets have reached a certain size, and so the layers can arise only when orbital forcing has driven the system to an unstable mode. Tracing the deposits to the earliest examples and looking for changes over the duration of the ice ages will clearly be useful. With the quick, nondestructive core-logging techniques now in use and the distinctive mineralogy of the layers, this should not be such a tall order. □

Eystein Jansen is in the Department of Geology, University of Bergen, Allegaten 41, N-5007 Bergen, Norway.

1. Bond, G. *et al.* *Nature* **360**, 245–249 (1992).
2. Heinrich, H. *Quat. Res.* **29**, 142–145 (1988).
3. Ruddiman, W. *Geol. Soc. Am. Bull.* **88**, 1813–1827 (1977).
4. Lehman, S. & Keigwin, L. D. *Nature* **356**, 757–762 (1992).
5. Karpuz, N. K. & Jansen, E. *Paleoceanography* **7**, 499–520 (1992).
6. Johnsen, S. J. *et al.* *Nature* **359**, 311–313 (1992).
7. Mangerud, J. *et al.* *Quat. Res.* **38**, 1–31 (1992).
8. Sancetta, C. *Nature* **360**, 249–251 (1992).

EVOLUTIONARY BIOLOGY

Strife among siblings

H. C. J. Godfray

PARASITOID wasps lay their eggs on or in the bodies of other insects, which act as food for the developing larvae. Polyembryonic species have eggs which divide asexually to produce a brood of genetically identical siblings that feed together on the same host individual. In 1907, the great Italian entomologist Filippo Silvestri¹ noticed that polyembryonic broods of some wasps are composed of two types of larvae: typical parasitoid larvae that develop into adult wasps, and larger precocious individuals that move around in the host haemocoel but die prematurely before pupation (see figure). Silvestri suggested that the precocious larvae macerate the host to facilitate feeding by their siblings, but most entomologists have accepted Cruz's² explanation, published eleven years ago in this journal, that the precocious larvae attack and kill other parasitoid eggs and larvae, so protecting their brood-mates from resource competition.

On page 254 of this issue³, Grbić and colleagues now argue that the function of precocious larvae is both more complex and more sinister than that. Precocious larvae tend to be females and attack clones of male larvae, their biological brothers. The presence of female but not male precocious larvae reflects the evolutionary resolution of an asymmetric genetic rivalry between male and female clones.

Polyembryonic reproduction is wide-

spread but uncommon in the animal kingdom from flatworms to armadillos. Among parasitoids it has evolved on four occasions, but precocious larvae are known only from wasps in the large chalcid family Encyrtidae. These insects oviposit into the eggs of moths and other insects, but their larvae do not begin to consume the host until it has grown into a mature caterpillar which can support up to several thousand individuals. Grbić *et al.* worked with perhaps the best known polyembryonic wasp species, *Copidosoma floridanum*, which attacks the eggs of noctuid moths.

Parasitoid wasps are members of the insect order Hymenoptera and so possess a haplodiploid genetic system, females developing from fertilized eggs and males from unfertilized eggs. The act of fertilization causes a noticeable pause in the sequence of egg-laying behaviours, and by careful observation of ovipositing wasps it is possible not only to count the number of eggs a wasp lays on a host but also to note their sex. Female *C. floridanum* either oviposit one or a pair of their own eggs into the egg of their host. If just a single egg is laid, it divides asexually to produce a brood of about 1,000–1,400 wasps, all of the same sex. When two eggs are laid, as is often the case, Strand⁴ found that one is invariably female and the other male. Again about 1,400 wasps emerge from the dead host, but of these about 1,200

Waning star

THE young stars of the Hyades cluster were found, in observation ten years ago by the Einstein satellite, to have X-ray luminosities up to 50 times higher than that of the Sun, a result presumed to indicate higher magnetic and chromospheric activity. Now the more sensitive ROSAT X-ray satellite has surveyed the cluster again, and finds some curious changes (R. A. Stern *et al.* *Astrophys. J.* **399**, L159–L162; 1992). Some stars which Einstein found to be similar to one another in X-ray brightness now appear quite different, and in one case a star that was easily detected ten years ago is not seen at all. The Hyades stars are not just more active than the Sun, but apparently more variable too, going through cycles like the 11-year solar cycle and perhaps, in the one case, entering into a quiescent phase like the seventeenth-century Maunder Minimum, when no sunspots were seen for decades, and the world experienced the little ice age.

One step back

WHY does actin hydrolyse ATP when it polymerizes? What does it need the nucleotide for anyway? Two years ago a paper in *Nature* offered an answer. It was that the elastic properties of filaments of ADP-actin and ATP-actin differ, the former storing elastic energy for use in later processes. But now T. D. Pollard and his colleagues (*J. biol. Chem.* **267**, 20339–20345; 1992) assert that this attractive scheme is a figment due to the instability of ADP-actin: if the protein is treated with proper respect, they say, the differences in elasticity and viscosity of the two types of filament vanishes. Their answer to the nucleotide enigma is that conversion of ATP- to ADP-actin allows for more rapid depolymerization of the filaments when the cell no longer has need of them.

First blood

INCISOR morphology is at the cutting edge of an explanation of why vampire bats are found only in the Neotropics (M. B. Fenton *Biol. J. Linn. Soc.* **47**, 161–171; 1992). Blood-feeding could have developed from feeding on maggots in wounds, an idea that assumes a plentiful supply of large, accident-prone mammals and birds. From this to wounding with teeth would be a small step. But why just in the Neotropics? Heritage is the key, says Fenton. The three genera of vampire all belong to the family Phyllostomidae, a diverse group confined to the New World. Phyllostomids combine catholic tastes with large, cutting incisors, whereas most other bats have very small incisors. The capacity to make new wounds may explain the dietary shift from blowflies to blood, and the absence of vampires in Transylvania.

are female and only 200 male.

It makes good evolutionary sense that mixed broods contain fewer males than females. Suppose that all the adult female wasps hatching from the caterpillar are mated by their brothers, and that the mating opportunities for males away from the brood are insignificant. The clone of male larvae will be selected to multiply to produce just enough individuals to mate with their sisters. What is more, natural selection will operate in exactly the same way on any

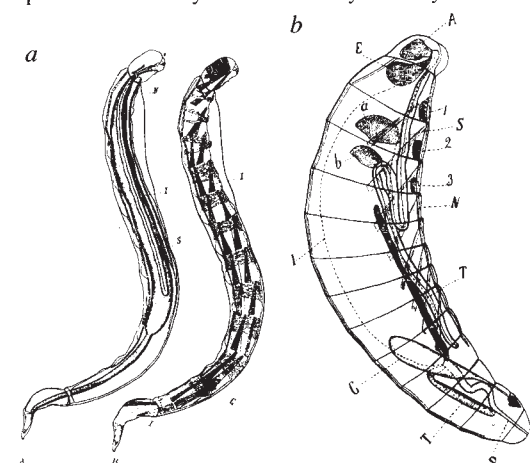
expectations, precocious larvae are almost exclusively female. Second, while female clones develop freely in the host thorax, male clones are hidden in fat tissue in the host abdomen. Moreover, *in vitro* the female precocious larvae attack developing male clones, but not female clones or host tissue. The function, or at least one of the functions, of precocious larvae thus seems to be the manipulation of brood sex ratio in favour of the female clone.

Whether the activities of the precocious larvae mean that the evolutionary conflict is decided in favour of the female clone is still not clear. A detailed quantitative comparison of the observed outcome with the optimum brood size and sex ratio from the point of view of the two clones is required to resolve this question. It is possible that proliferation within the male clone, partly protected in the fat tissue, manages to balance losses from female attack so that the outcome is in fact nearer the male optimum.

In general, the study of evolutionary conflicts within the family has two stages. The first stage (usually the easier) is the delimitation of the battleground, typically by comparing the evolutionary optima for the two combatants. The second is the prediction of the resolution of the conflict, which is much harder as it normally requires a detailed understanding of the coevolutionary interactions between the two participants. Because of their experimental tractability, polyembryonic wasps may provide an ideal model system to study the resolution of an evolutionary conflict.

In *The Selfish Gene*, Dawkins⁵ notes that no feats of "heroic self-sacrifice" are recorded from broods of polyembryonic nine-banded armadillos but that "it would be well worth somebody's while going to South America to have a look". An entomological chauvinist would point out there is no need for that, as both heroic self-sacrifice and the most devilish of selfish genery can be found in the polyembryonic wasps that parasitize the caterpillars in most people's gardens. □

H. C. J. Godfray is in the Department of Biology and NERC Centre for Population Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK.



Precocious (left) and normal (right) larvae of the polyembryonic wasp *Copidosoma (=Litomastix) truncatella* (two figures not to scale). The figures are taken from Silvestri's article¹, in which he described for the first time the larval polymorphism and the fate of the precocious larvae.

genes expressed in the female larvae or the parent that influence sex ratio. In other words there are no genetic conflicts of interest between parent, sons and daughters.

But if males are able to obtain a significant number of matings away from their natal brood, conflicts of interests rear their heads. Natural selection acting on genes expressed in the male larvae will favour a larger male clone size to capitalize on the opportunities to mate with other females in the environment. However, natural selection acting on genes expressed in the female clone will favour a reduction in the size of the male clone to reduce competition for the limited resources that constitute the host. The difference in evolutionary optima is heightened by a genetic asymmetry common to all haplodiploid organisms. There is a greater probability that a gene present in a male is also present in his sister, than that a gene present in a female is also present in her brother. In other words, a male cares more for his sister than a female does for her brother.

It is against this background of potential evolutionary conflict between male and female clones that the experiments of Grbić *et al.* are particularly interesting. They found two sexual asymmetries in development. First, and against all

DAEDALUS

Metal blowing

METALS and their alloys are usually crystalline. Ultra-fast cooling can produce amorphous, glassy metal in tiny amounts, much prized for its distinctive properties. Daedalus now proposes to make it in bulk. His scheme depends on the fact that the electrical resistance of almost every metal rises on melting.

Imagine, he says, a molten metal carrying a current; and imagine a small crystal forming within it. Its lower resistance will divert the current preferentially through itself, forming a local short-circuit. The resulting concentration of heat could melt the crystal again. So a suitable current could prevent a metal or alloy from crystallizing as it cooled. The tiny crystal nuclei, essential to start the process, would be obliterated as fast as they formed. Ultimately, the metal would get too cold to crystallize at a detectable rate. The current could then be turned off and the product quenched to room temperature. A supercooled liquid, a true metallic glass, would result.

DREADCO's metallurgists are now melting all sorts of alloys in a modified induction furnace, and slowly reducing (but not shutting off) the current. With proper control, the current-carrying metal should solidify in glassy form. The resulting alloys should have all sorts of interesting new mechanical, chemical and magnetic properties.

For a start, they might have a very wide range of useful compositions. The craziest random assortment of scrap metals, quite incompatible if melted into a crystalline alloy, should solidify in the induction current-field as a perfectly homogeneous and useful metallic glass. Furthermore, a really complicated mixture of metals should have a very low effective melting point. As a glassy alloy, it should soften to a viscous liquid which could be shaped, moulded or even blown at not much above room temperature.

Once metal can be blown like glass, hollow metal objects from teapots and kettles to car bodies and airframes could be blow-moulded in one operation, simply by scaling up standard glass-working techniques. At the highest moulding temperatures, the metal might begin to crystallize again; but that could be an advantage too. If the hot product were 'seeded' at one point by a single crystal at the desired orientation, the crystallization front would spread uniformly through the whole mass. The resulting structure, however complex, would consist of one single grainless metallic crystal. DREADCO metallurgists are divided as to whether it would be immensely strong or absurdly flexible. Either way, exciting new engineering techniques should open up.

David Jones

1. Silvestri, F. *Boll. Lab. Zool. gen. agr. R. Scuola Agric. Portici* **1**, 17-64 (1907).
2. Cruz, Y. P. *Nature* **294**, 446-447 (1981).
3. Grbić, M., Ode, P. J. & Strand, M. R. *Nature* **360**, 254-256 (1992).
4. Strand, M. R. *J. Insect Behav.* **2**, 355-369 (1989).
5. Dawkins, R. *The Selfish Gene* (Oxford University Press, 1976).

Last resting place of a thylacine

SIR — The published photographs^{1,2} of lairs and dens of the thylacine, Tasmanian tiger or Tasmanian wolf (*Thylacinus cynocephalus*) are all associated with unsubstantiated observations and sightings of animals in the wild; the last specimen in captivity died in the Beaumaris Zoo, Hobart in 1936. The nature and purpose of the lairs is unknown: for example, the size of the animal's home range and whether the resting places chosen during the day were casual and transient, or semi-

formed part of the director's report to the 1,127th meeting of the council in 1903: "I visited Tasmania in December, . . . and during that time was enabled to travel round a portion of the coast via Scottsdale, . . . and was enabled to visit many small settlements, such as Pioneer, Derby, Cullenswood, Avoca etc., in country where Tasmanian Wolves and Devils are found. I gave lectures with the aid of my magic lantern at each place and well advertised our Gardens, besides giving two public lectures in Launceston,

so that if any of the animals mentioned are caught in these districts, they will come to our Gardens. I was also enabled to bring back with me a pair of Tas. Devils, 2 Black opossums, besides obtaining a Tas. Wolf"³. While in the Avoca district, Le Souef stayed with family friends, the Franks, at their property Meadstone, on the upper reaches of the St Paul's River. It was here, while Le Souef was visiting, that a thylacine was captured from the lair in the figure, and the station hands involved were paid £7 for their effort³.

The lair itself appears as a shallow cave in the St Paul's River Triassic sandstone, with some nesting material on the floor, much of which appears to be made of fern fronds similar to those of the specimen of *Polystichum proliferum* growing outside the lair. The photograph reflects the most commonly described location for thylacine lairs in the literature, that of their placement within caves, rocky hollows and rocky dens^{4,5}.

Le Souef returned to Melbourne with his thylacine on 24 December 1902 and wrote to Krueger on 1 January 1903, offering him the animals he requested, at the prices he had suggested, including: "one Wolf (Tasmanian) . . . price twenty pounds". This thylacine, captured from its lair and despatched from Melbourne Zoo in 1903, represents the first of several thylacines held in captivity in the Antwerp Zoological Gardens; another was obtained in 1913 from the Beaumaris Zoo in Hobart, and Moeller also notes⁶, in reference to the Royal

Society and Regent's Park Zoo, that "Antwerp . . . received several Tasmanian wolves, via London".

In conclusion, I suggest that the historical records associated with this particular specimen adequately validate its existence. Hence, for the first time, a photograph of the last resting place of a thylacine in the wild has been authenticated.

R. N. Paddle

Christ Campus,
Australian Catholic University,
PO Box 213, Oakleigh,
Victoria 3166, Australia

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The thylacine lair in the Triassic sandstone of the upper reaches of St Paul's River, Tasmania. (Photograph from the La Trobe Collection, State Library of Victoria.)

permanent to permanent, is open to question. It is not clear, therefore, whether the recent photographs^{1,2} are really of thylacine lairs.

My examination of the historical records of the Royal Melbourne Zoological and Acclimatisation Society (RMZAS) has revealed the first unambiguous photograph of a lair from which a thylacine was caught, and although no photograph has been preserved of the specimen, sufficient records of its existence and eventual fate have been found to validate the status of the photograph (see figure).

At the 1,126th meeting of the council of RMZAS in 1902, a letter was read from F. Krueger, director of the Zoological Gardens in Antwerp, who wished to purchase certain Tasmanian animals at a generous price³. The Melbourne Zoo's director, W.H.D. Le Souef, noted "as we have none to spare and they are difficult to procure, I would suggest . . . that I go to Tasmania tomorrow for about a fortnight and visit the various centres where these animals are found, and propose taking my magic lantern with me, so as to interest the people I visit in our Gardens and requirements and we should make a good profit . . . I anticipate the total expenses will be under fifteen pounds". The RMZAS council readily agreed to his proposal³.

Le Souef's activities in Tasmania

Universal cellular tropism?

SIR — Although human immunodeficiency virus-1 (HIV-1) has been demonstrated to infect an ever-increasing array of human cell types, isolation and localization of the virus *in vivo* suggests that cells of the lympho-reticular system are predominantly affected. Understanding the viral-specific determinants, such as the viral envelope, which govern the susceptibility of various lympho-reticular cells to HIV-1 (a process called cellular tropism) and subsequent viral production during the acute and chronic stages of infection is critical to the design of vaccines and immunotherapeutics. Although not ideal, the *in vitro* propagation of various human cells either of primary origin or as immortalized cell lines provides a means to study those structural elements of the viral envelope responsible for tropism.

Various studies have described the molecular determination of cellular tropism of HIV-1 as residing in the viral envelope¹⁻³. More specifically, cellular tropism was mapped by the authors of these studies to the envelope region which was found to include the third hypervariable domain (V3), an epitope known previously to be involved in neutralization, cell fusion, viral infectivity, and both T- and B-cell epitope recognition (reviewed in ref. 4). In these studies, three sources of human cells were used to map the envelope regions responsible for tropism: PHA-stimulated peripheral blood mononuclear cells (PBMCs), which constitute a mixed primary T4 helper cell population; various immortalized human T4 cell lines; and primary macrophages derived from the PBMC pool. Although significant differences in infectivity and replication on the T-cell lines and macrophages were noted with viruses containing chimaeric recombinant envelope, no differences were observed for infectivity and replication kinetics in the PHA-stimulated PBMCs (Table 1 of ref 1; Figs

1. Park, A. *Austr. Geogr.* **1**, 66-83 (1986).
2. Guiler, E. R. *The Tasmanian Tiger in Pictures* (St David's Park, Hobart, 1991).
3. Royal Melbourne Zoological and Acclimatisation Society *Minute Book No. 8. 1893-1903* (Archives of the RMZAS, Public Records Office, Melbourne).
4. Guiler, E. R. *Thylacine: The tragedy of the Tasmanian tiger* (Oxford Univ. Press, Melbourne, 1985).
5. Smith, S. *The Tasmanian Tiger - 1980* (National Parks and Wildlife Service, Hobart, 1981).
6. Moeller, H. F. in *Grzimek's Encyclopedia of Mammals* Vol. 1. (McGraw-Hill, New York, 1990).

1 and 2 of ref. 2; and Table 2 of ref. 3). This distinct lack of tropism (or the more universal permissiveness of PBMCs for HIV-1 regardless of the specific envelope regions) was not discussed in these papers.

We feel that this result is important and deserves additional comment. It suggests that viral envelope elements that otherwise abolish infectivity of cell lines are tolerated to a significant degree in human primary PBMCs, thereby suggesting fundamental differences of viral entry at the cellular level for these cells. In support of this concept of 'universal tropism', 66 isolates were found to replicate primary PBMCs in a similar fashion regardless of the clinical stage of the disease⁵. Furthermore, this fundamental difference, which we describe as "enhanced permissivity", appears also to be related to a more soluble CD4 (sCD4)-resistant state. In recent experiments, when PBMCs instead of transformed human T4 cell lines were used in infectivity titration experiments designed to study the decay rate of HIV-1, a doubling of the infectivity half-life, from 36 to 53 hours, was observed⁶. The sensitivity to sCD4 blocking of viral infection was also reduced by about 20-fold.

These cellular factors operating in conjunction with viral envelope factors could begin to explain the growing body of experimental evidence indicating the greater degree of resistance in sCD4 blocking and neutralization by antibodies observed for both primary isolates and laboratory strains. As we consider the issue of tropism and attempt to

model more predictive *in vitro* systems of viral infectivity, an appreciation of both viral and cellular variables should improve our understanding and lead to more effective initial immunoprophylactic and therapeutic approaches. These results⁶, coupled to the more 'universal tropism' exhibited in the previous studies¹⁻³, imply that viral absorption, binding and penetration of HIV-1 is fundamentally a more efficient and permissive process in human PBMCs. The greater resistance to blocking and infection by sCD4 and neutralizing antibodies, in addition to the prolonged infectivity exhibited during the decay studies, suggests that other cell surface factors are contributing to viral entry.

Peter Nara

Wen-Po Tsai

Hsiang-Fu Kung

Anton Minassian

Robert Garrity

National Cancer Institute,

Frederick,

Maryland 21702, USA

Jaap Goudsmit

Gustaaf Rimmelzwaan

Virology Department,

Academic Medical Centre,

Meibergdreef 15 1105 AZ,

Amsterdam,

The Netherlands

1. O'Brien, W. A. *et al.* *Nature* **348**, 69-73 (1990).
2. Shioda, M., Levy, J. A. & Cheng-Mayer, C. *Nature* **349**, 167-169 (1991).
3. Hwang, S. S., Boyle, T. J., Lyster, H. K. & Cullen, B. R. *Science* **253**, 71-74 (1991).
4. Nara, P. L., Garrity, R. & Goudsmit, J. *FASEB J.* **5**, 2437-2455 (1991).
5. Lu, W. & Andrien, J. J. *Virology* **66**, 334-340 (1992).
6. Layne, S. *et al.* *Virology* **189**, 695-714 (1992).

Even larger organisms

SIR — Extremism, in the pursuit of science, is apparently no vice. We wish to contribute to the discussions about the largest living organisms¹ which were stimulated by the nomination² of *Armillaria bulbosa* as among the largest of living organisms. Two measures of size were reported: mass at 10,000 kg and area at 15 ha for this fungus. Two other rather more famous entities have traditionally headed the list of living behemoths: the giant sequoia at just less than 2,000,000 kg (ref. 3) and the blue whale at a paltry 180,000 kg.

In the interests of big science, we wish to nominate the quaking aspen, *Populus tremuloides*, as the largest living organism whether measured by area or mass. At least one clone in the western United States cited in ref. 4 covered 43 ha, nearly three times the area reported for *Armillaria*. This aspen clone was estimated to contain 47,000 stems. An average individual aspen shoot (stem plus leaves and branches) would easily exceed 100 kg in mass while the accom-

panying root system would probably exceed 30 kg. Combination of these rather conservative figures produces an estimate of approximately 6,000,000 kg mass for that particular aspen clone, a number 60 times larger than the mass of the *Armillaria* thallus and more than triple that of the largest giant sequoia, the species most widely acknowledged as the mass champion of all plants². Quaking aspen, the most widespread tree species in North America, can now take its rightful place as an acknowledged giant among giants.

Michael C. Grant

Jeffrey B. Mitton

Yan. B. Linhart

Department of Environmental, Population

and Organismic Biology,

University of Colorado,

Boulder, Colorado 80309, USA

1. Brasier, C. M. *Nature* **356**, 382-383 (1992).
2. Smith, M. L., Bruhn, J. N. & Anderson, J. B. *Nature* **356**, 428-431 (1992).
3. Elias, T. S. *Field Guide to North American Trees* (Grolier Book Clubs, Danbury, Connecticut, 1989).
4. Barnes, B. V. *Forest Sci.* **21**, 319-328 (1975).

Motor neuron disease model

SIR — M. Sendtner *et al.* (*Nature* **358**, 502-504; 1992) report that in a mouse mutant with progressive motor neuronopathy, ciliary neurotrophic factor (CNTF) released by implants of genetically modified fibroblasts increases the survival of facial motor neurons and prolongs the lifespan of the diseased animals. However, we became puzzled about the animal model that they used. In particular, the results in their Table 1 show a relatively modest loss of facial motor neurons at a time when the animals are terminally ill. Can such a small degree of motor neuron loss account for the death of these animals?

We turned to the original description of the mutant used (H. Schmalbruch *et al.* *J. Neuropath. exp. Neurol.* **50**, 192-204; 1991). To our surprise, the authors of this paper concluded that even at the terminal stage of the disease there was no loss of motor neurons. "The cervical and lumbar spinal cord of diseased mice in sections stained with cresyl violet appeared normal. It was not possible to identify unmarked sections from normal or diseased mice." In another part of the paper the authors say: "All nuclei of cranial nerves were present and did not reveal qualitative or obvious quantitative changes." This is in sharp contrast to the statement by Sendtner *et al.* that "The motor neurons of *pnn/pnn* mice first undergo a reduction in cell size, then chromatolysis and finally cell death."

These discrepancies make us wonder whether Sendtner *et al.* used a different mutant, or whether the colony has changed. But if the original description of the disease is correct and the disease affects predominantly motor nerve fibres and not cell bodies as well as nerve fibres, then the interpretation of the results of Sendtner *et al.* is entirely different from that proposed by them, that is, that CNTF rescues motor neurons in this mouse mutant which, according to the previous description, suffers from no loss of motor neurons.

The results are no less interesting if genetically manipulated cell implants that release CNTF protect nerve fibres from dying back rather than rescuing motor neurons. However, the theoretical and clinical implications are entirely different.

Gerta Vrbová

Linda Greensmith

Katarzyna Sieradzka

Department of Anatomy and

Developmental Biology,

University College,

Gower Street,

London WC1E 6BT, UK

The truth about physics

J. L. Heilbron

The Creative Moment: How Science Made Itself Alien to Modern Culture. By Joseph Schwartz. Jonathan Cape/HarperCollins. Pp. 252. £16.99, \$25.

MANY will share Schwartz's indignation at our greedy, ignorant despoliation of our planet. Fewer, perhaps, will accept his diagnosis that the fatal flaw in our civilization is the refusal of almost all of us to learn enough about our physical world to understand the environmental consequences of our greed or the beauty of the science our acquisitiveness exploits. Fewer still will follow Schwartz in identifying misunderstanding of the theory of relativity as an important cause of the decline of the West.

The route by which he arrives at this last insight is as arresting as the insight itself. According to Schwartz, Galileo and Newton started the trouble by couching physics in terms of mathematics. They did so from no necessity immanent in the subject matter but to ensure the survival of physics in a hostile political and religious environment. "If physics were to survive in a socially divided Europe, it had to be well hidden". The message of Galileo was "keep it quiet, keep it obscure, and keep it mathematical". Newton's *Principia*, according to Schwartz, can easily be decoded to reveal its physical content apart from its mathematics. The example he offers, which reasons from "the observed acceleration of the planet Mercury in its orbit", is not reassuring. Mercury's acceleration is not an observation but an imputation, authorized and confirmed by the mathematics of the theory of gravity.

The culmination of nineteenth-century physics, according to Schwartz, is the theory of relativity. Once again, mathematics veiled physics from the people. Once again, Schwartz guts the physics to make his point. The entire physical content of the special theory of relativity, he says, is that all physical interactions require time. But as the case of the propagation of sound waves in air suggests, it is not possible to deduce from the principle of action in time that there exists a maximum permissible velocity that has the same value for everyone in the Universe. The deduction requires additional postulates, and the demonstration of their counter-intuitive consequences requires mathematics.

Just as physicists of Newton's time missed the chance to make their subject popular, so in Schwartz's rendition of history their descendants in the 1920s, failing to take advantage of the popular enthusiasm for Einstein, managed to

make the beautiful and transparent theory of relativity a byword for the difficult and obscure. In this we may hear the cry of the proselytizing pedagogue. Schwartz is the author of a book intended to make Einstein's theories easy. He must often have encountered the strong resistance of humankind to mathematical reasoning, even when covert, and people's general distaste for mathematicians. Benjamin Franklin's tribute to his friend Thomas Godfrey may stand for many expressions of this antipathy. Godfrey was not "a pleasing companion," Franklin wrote, "as like most mathematicians I have met with, he expected unusual precision in everything said, or was forever denying or distinguishing upon trifles to the disturbance of all conversation." Because of its mathematics, physics will not appeal to all tastes. It can be popularized qualitatively and usefully; but the result is not physics.

Much of the rest of Schwartz's argu-

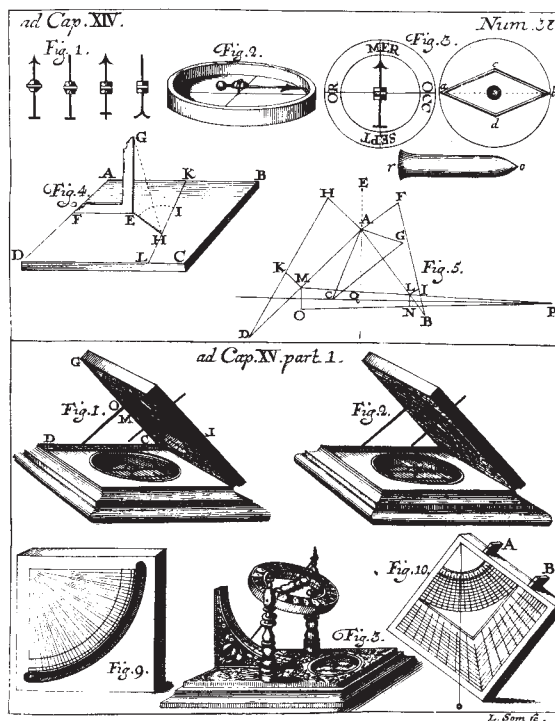
ment concerns the separation of the cultures of production and consumption. With the physicists' conspicuous failure to popularize their subject, consumers lost track of the physical consequences of consumerism. Physics itself split into a theoretical part, increasingly removed from nature, and an experimental part, closer to real things and consequently further from the pinnacle of arrogance where the élite of the élite, the theorists, prance without advancing. The failure of the physicists during the 1920s to understand quantum mechanics coincided with the failure of popular political revolutions in Central Europe and also with the failure of new art-forms, such as cubism and music on the twelve-tone scale, to establish themselves. The physicists did succeed in one thing, however: they attracted the support of the Rockefeller Foundation. That was a mistake. Physicists supported by the foundation erred in thinking that they retained control of their work. "The political turning point had passed. Alamogordo was now very near."

Having come so far, Schwartz attacks a menagerie of *bêtes noires*: General Groves and the Manhattan Project, for duping physicists to make a bomb to bully the Soviet Union; the Rockefeller Foundation again, for converting molecular biology into a mechanistic science;

Time on one's hands

During the sixteenth and seventeenth centuries, diptych sundials became popular devices for telling the time. Most diptych dials have a string gnomon (shadow-caster) which stretches at an angle (determined by the latitude of the observer) between the two tablets. Magnetic compasses for correctly orientating the dials were probably first added before the fifteenth century. This illustration from a work by Johannes Gaupp (1711) shows various styles of compass needles and bowls; it is reproduced in *Ivory Diptych Sundials, 1570–1750*, a beautifully illustrated catalogue of Harvard University's collection of 82 of these timepieces.

The collection, which is one of the largest in the world, consists mostly of designs from Nuremberg (the first city in Europe to have an organized trade of compass-makers), Paris and Dieppe. The book is distributed by Harvard University Press, £39.95.



the scientific establishment, for insisting that the human immunodeficiency virus causes AIDS; the inventors of IQ tests and others who teach that intelligence is hereditary; promoters of the Superconducting Super Collider; managers ignorant of production processes; heads of corporations who put profit before safety; NASA politicians who oversell their products; everyone who mystifies relativity; and post-modern literary critics.

Most of what Schwartz attacks deserves censure. The greed and self-

centredness of Western civilization and the Malthusian behaviour of developing societies menace the future of all rational animals. But physics is neither the cause nor the solution of our predicament. Its history does not make a very persuasive symbol of the decline of our civilization. □

J. L. Heilbron is at the Office for History of Science and Technology, University of California, Berkeley, California 94720, USA.

Macromolecules and medicine

Tom Blundell

Protein Structure: New Approaches to Disease and Therapy. By Max Perutz. W. H. Freeman: 1992. Pp. 326. \$39.95, £32.95 (hbk); \$29.95, £21.95 (pbk).

WHEN in 1972 we learnt that the Chinese had extended their research on insulin from a complete chemical synthesis to a successful crystal-structure analysis, it was rumoured that Mao Tse-Tung must have diabetes. Why else would they have maintained this research when almost all other biological projects had been terminated by the cultural revolution? But the truth was otherwise. Scientists in Shanghai who had previously studied in Cambridge alongside Fred Sanger, had been challenged by the Great Leap Forward and had read Engels: "If you know the formula, then synthesize it!" As insulin was the only protein sequence then known, they proposed to synthesize it. Chou En-Lai was impressed, and insulin became central to the new Chinese science. Later, encouraged by Dorothy Hodgkin, they crystallized their synthetic protein and became fascinated by the relation of structure and function.

In *Protein Structure*, Max Perutz describes many other equally intriguing tales, showing how results of great medical significance have evolved from quite surprising sources. In 1934, J. D. Bernal and D. Crowfoot (as Dorothy Hodgkin then was) reported the X-ray diffraction from protein crystals for the first time. The protein was pepsin. Fifty-five years later, crystal structures were defined for two other family members: human immunodeficiency virus proteinase, central to the design of AIDS antiviral drugs, and renin, equally important in the design of antihypertensive drugs. Inhibitors of both enzymes had been previously sought by many pharmaceutical companies, and the structures of inhibitor complexes of several pepsins had already provided much useful assistance in design.

In a similar way, Perutz tells us that he came to England in 1937 because "the structure of proteins seemed to be the most fundamental unsolved problem in biochemistry and X-ray crystallography the only method of solving it". He chose haemoglobin because it was "[t]he easiest protein to crystallise". However, his book shows that haemoglobin has been of much interest as a drug-receptor model and as a basis for understanding the molecular pathology of human haemoglobin.

The general importance of protein structure to medicine was recognized at an early stage. When Hodgkin obtained a diffraction pattern from insulin crystals in 1935, she was well aware of the role of insulin in diabetes — especially as insulin crystallization was a route to purification and insulin crystals a useful form for administration of the hormone. Certainly Perutz himself was aware of the importance of protein structure when, with the support of Lawrence Bragg, he persuaded Harold Himsworth, secretary of the British Medical Research Council, to fund his research at the Cavendish Laboratory in Cambridge. And so the theme of the book: was this research justified?

Perutz answers in the affirmative with an enthusiastic account, mainly of progress during the past five years. This is a *tour de force*. Perutz selects his examples carefully to illustrate general themes, but in each case brings out the importance of the protein structure to drug design. In "How Proteins Recognize Each Other" he skilfully uses the immunoglobulin fold as an example. He describes how the structural motif is used in antibodies, the major histocompatibility complex, the T-cell receptor and cell adhesion molecules, and how nature has used it in different ways in molecular recognition.

Indeed, for Perutz it was the successful engineering of an antibody against T cells that provided the real justification for his confidence 40 years earlier. He describes how T. A. Waldmann and G.

Winter humanized a rat anti-T-cell antibody. First, oligonucleotides that code for the regions defining the specificity (the CDRs) of the rat antibody were grafted onto the genes coding for the heavy and light chains of two different human myeloma proteins of known structure. Then the nucleotides coding for the C-terminal domains of another myeloma protein, known to activate complement, were grafted onto the gene for the heavy chain. But the resulting antibody was inactive. The solution to the problem came from known three-dimensional structures. A serine in the human protein was exchanged for phenylalanine on the basis that it was needed to prop up one of the CDRs. The resulting protein was then active and was used successfully to treat patients with non-Hodgkin lymphomas, for which other treatments had failed.

Much of the book is implicitly about technology transfer. It is evident that the new knowledge about protein structure must be the subject of continuing discussion between protein crystallographer and drug designer, who will often be in a biotechnology or pharmaceutical house. The products of first designs are brought back by the chemists for further structural studies, and the design cycle is repeated, often many times. Perutz shows how the ability to engineer changes in the protein can also provide complementary information about binding specificity; the drug design and protein engineering cycles must be ridden in tandem. Moreover, the applied work has spin-offs for our understanding of fundamental processes. There can be no better argument for keeping basic and applied research closely integrated.

Throughout the book, Perutz is remarkably up to date; he includes many examples that have been published in the past year. This is particularly evident in his discussion of how proteins recognize genes and in the section on cytokines, growth and differentiation factors. Even though he retired some time ago, Perutz is probably the best informed of protein crystallographers, with a very good sense of where the excitement is. He has avoided the purely historical account, which must have been seductive, and has clearly been on the telephone to investigators at the frontiers. In 1964 my tutor at Oxford advised me not to go into protein crystallography as Perutz and J. C. Kendrew had made the major breakthrough. The rapid progress over the past years, now described by Perutz, is ample reward for my decision!

Characteristically, Perutz does not avoid the principal challenge of outlining to an audience, probably largely consisting of non-physicists, the techniques and theory of protein crystallography. His introductory chapter is an example to all

teachers. I remember how, 20 years ago, I set out to write a book on protein crystallography for biologists, but gave it up as too difficult. Instead I wrote, with Louise Johnson, a book for specialists of more than 600 pages. More thoughtfully, and much more expertly, Perutz gives an introduction to protein crystallography and NMR in 39 pages; his description of his own guru Bragg as "artistic, with a lucid, visual grasp of physical phenomena" applies equally well to him. The introduction is indeed "Diffraction Without Tears", and I recommend it to all teachers in biology who have the unenviable job of describing the method to rather nonmathematical students.

Perutz likens the diffraction pattern of a protein crystal to a hieroglyphic without a key, but recognizes that his analogy limps because, once the key had been found, the Rosetta stone, for example, was deciphered word for word. On the other hand, X-ray crystallographers toil away at their structures, often for many years, with no hint of any features until the final dramatic moment. But Perutz demonstrates easily that each structure contains a wealth of information when that moment does arrive. From his own studies he derived a useful set of generalizations about interactions between drugs and proteins. The affinity of a drug increases with the number of water molecules displaced on binding the protein, and with decreasing loss of rotational entropy. Small changes of drug stereochemistry can exploit new binding sites on the protein, and binding at very different sites can influence the allosteric equilibrium in a useful way. On the question of preformed sites versus induced fit, Perutz is eloquent. He compares the alternatives to those of the rich man who has a suit made to measure and the poor man who buys one off the shelf, but remarks that in the case of antigens, antibodies are taken off the peg, but "small alterations do improve the fit".

This is a book for the beginner and for the specialist alike. It is a beautifully written survey of an important field. Its illustrations lack the colour and the uniform artistry of other recent texts, but that is made up for by the enthusiasm and the knowledge that we can share in reading it. □

Tom Blundell is in the Structural Molecular Biology Unit, Imperial Cancer Research Fund, Birkbeck College, Malet Street, London WC1E 7HX, and at the Agricultural and Food Research Council, Polaris House, Swindon SN2 1UH, UK.

■ The second edition of *Proteins: Structures and Molecular Properties* by Thomas E. Creighton is also newly published by W. H. Freeman, price £22.95.

Sense and sensibility

Richard G. Vile

The Transformed Cell: Unlocking the Mysteries of Cancer. By Steven A. Rosenberg and John M. Barry. Putnam/Chapmans: 1992. Pp. 353. \$24.95, £15.99.

"It is every immunologist's dream to use the immune system to cure cancer. It is every immunologist's nightmare that there is no such thing as an immune response to cancer in humans." This quote comes early on in Steven Rosenberg's highly readable and thought-provoking autobiography and epitomizes



Rosenberg — hopping between laboratory and clinic.

his lifelong crusade to bring the fruits of the laboratory into the cancer clinic. By the end of the book, it seems clear that his particular nightmare is over, but his dream is still far from realization.

This is the story of what is possible when intense personal motivation is coupled to an important question and an excellent opportunity. The motivation arose from a young man's hatred of needless pain, a hatred stemming from the deaths of many of his relatives in Nazi concentration camps. Rosenberg became obsessed with doing something that "very, very few" can achieve — high standard research that he could then apply to patients suffering from cancer, a disease he describes as a holocaust. But his path was far from easy. For example, when the welcome attentions of his future wife Alice O'Connell, the head nurse of the emergency room at the Brigham Hospital in Boston, threatened to distract him from his work, he briefly sought counselling from Harvard University's psychiatric service.

Opportunity presented itself with his meteoric rise at the age of 34 to the position of chief of surgery at the US National Cancer Institute. With an annual budget of several millions of dollars at his disposal, he was now able to address his important question.

During his clinical training, Rosenberg examined a patient and was surprised not to find any trace of the stomach cancer that should have killed the man 12 years before: "this man's body had cured cancer", he writes, "but how?" Rosenberg proceeds to give a clear description of a vast series of experiments designed to discover how the immune system might recognize its own cancer cells as 'foreign' and attack them as it would an invading pathogen. But this discipline of 'immunotherapy' was still viewed with immense scepticism by

many scientists. Nonetheless, Rosenberg set out to isolate immune cells that could recognize cancer cells, and with the help and insight of his wife he devised a five-point strategy by which the search would have to proceed if it were to lead to clinical treatment. Each major scientific advance — Rosenberg's "thunderbolt moments" — led to the ticking off of points, including the initial identification of immune cells that, when grown in the cytokine interleukin-2 (IL-2), can kill tumour cells non-specifically *in vitro*. These lymphokine-activated killer (LAK) cells, purified from

mice with tumours and readministered with IL-2, can cure metastatic disease. A "desperate search" for enough IL-2 was eventually solved using recombinant DNA technology.

LAK cells were eventually superseded with the discovery of more potent, tumour-specific T lymphocytes, named tumour-infiltrating lymphocytes (TIL), which can be purified from a patient's tumour, expanded *in vitro* and readministered to the patient to seek out and destroy tumour deposits. The work is still in progress. Rosenberg describes "the new direction" of research, in which potentially therapeutic genes are inserted into TIL for delivering high quantities of immune-stimulating or tumour-killing proteins locally to tumour cells — a goal that is beyond the reach of conventional chemotherapy.

Rosenberg stresses that he has never been merely engaged in an intellectual exercise. Every important finding at the bench has a corresponding application to patients, all of whom have failed to

respond to conventional therapies and are not expected to live more than 90 days. And it is here that we face the most controversial issue of Rosenberg's work. For although his laboratory studies have been exceptionally successful — he has written more than 500 scientific articles — the treatments have been fraught with disappointment and distress for both doctors and patients. All of the first 21 patients treated with LAK cells died of their cancers, and many of the patients who received IL-2-based therapies have experienced severe side-effects (IL-2 is highly toxic).

Indeed, there have been cases in which the treatment itself has killed. These deaths, which are at the usual frequency for experimental therapies for cancer, highlight Rosenberg's determination to succeed. He admits that it is this aggressive but necessary nature of his therapies that has attracted the most criticism from both the press and his peers: only patients with otherwise untreatable disease are eligible for the trials, and the new therapies have to excel if they are to make a good impression. The considerable physical pain of the treatments to the patients is honestly described, and Rosenberg does not shirk recalling the failures of his treatments and the personal torment they cause him.

Yet there have also been triumphs. The first patient to receive LAK cells and IL-2 in high doses is alive and well seven years after arriving in Rosenberg's clinic with a huge and spreading tumour. In other patients, tumours have shrunk but later re-appeared. About ten per cent of treated patients have seen their tumours vanish completely. Rosenberg argues that this minority "speaks powerfully" for his form of immunotherapy.

It sometimes seems paradoxical that the hatred of senseless pain that initially drove Rosenberg should result in such aggressive treatments. Yet the sensibility of Rosenberg to his patient's needs cannot be doubted; not only do the patients desperately want to be cured, but many also feel an urge to help in new trials that may eventually benefit others to come.

It may be the dream of many cancer scientists to have an office from which one can exit by one door to the laboratory and by the other to the ward, combined with the resources to carry ideas with ease from one to the other. Don't dream too hard though — the good sense behind Rosenberg's ideas and the punishing sensibility required to prove them are clearly difficult to reconcile. □

Richard G. Vile is at the Imperial Cancer Research Fund, PO Box 123, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.

Dynamics in the balance

John M. Wallace

Physics of Climate. By José P. Peixoto and Abraham H. Oort. *IOP/AIP: 1992. Pp. 520. £66, \$95 (hbk); £31.25, \$45 (pbk).*

THIS book is dedicated to the late Victor P. Starr, professor of meteorology at the Massachusetts Institute of Technology and the authors' mentor during the late 1950s and early 1960s. Starr's idea of obtaining a mathematically rigorous set of balance requirements for the conservation of mass, momentum and energy and of using it as a framework for diagnosing the complex workings of the atmospheric general circulation, has had a profound influence on the development of the field of climate modelling. Here Peixoto and Oort review the application of these methods to an assortment of physical parameters encompassed by the global climate system.

The authors address a well chosen subset of the issues. The book is divided into 17 chapters, most of which are fairly self-contained. Some would be suitable as chapters of a textbook, whereas others are more in the spirit of review articles. The first two chapters provide a clear definition of the climate system and a concise and readable discussion of fundamental concepts that provide a basis for a quantitative discussion of climate variability and its causes. Chapter 3 presents a highly compressed derivation of the governing equations. In Chapter 4, the authors describe how the atmospheric and oceanic circulation can be decomposed into time and zonal mean components and departures from them; this section constitutes essential background material for understanding many of the observational results presented in later chapters. Chapter 5 introduces the reader to climate data sets and data analysis techniques. Chapter 6 is a comprehensive and reliable treatment of radiative transfer in the atmosphere and the observed global radiative balance.

In Chapters 7–15, which are in a sense the core of the book, the authors review the three-dimensional structure of the principal components of the climate system and document the exchange processes at the atmosphere's lower boundary, the angular momentum balance, the hydrological cycle, the balance of total energy, the kinetic energy cycle and the generation of entropy. These chapters are densely illustrated with maps and vertical cross-sections; they include discussion of the balance requirements themselves and their implications for

climate variability. Oceans and cryosphere are incorporated into the discussion, as appropriate.

In Chapter 16, the authors present a readable and informative discussion of interannual and interdecadal climate variability. Of all the material in the book, I found it the richest in insights into how the climate system works. The final chapter provides an introduction to climate modelling and a review of some of the applications of climate models.

Most of the chapters should be readily accessible and useful to graduate students and postgraduates in the physical sciences who are seeking an understanding of the observed climate and the causes of climate variability. Chapter 4 and Chapters 11–14 will appeal to a more specialized audience of atmospheric and oceanic dynamicists and modellers. Those wanting an introduction to the governing equations may find Chapter 3 overly terse. More expansive treatments with examples and illustrations are available in J. R. Holton's *An Introduction to Dynamic Meteorology* (Academic, 3rd edn, 1992) and A. E. Gill's *Atmosphere–Ocean Dynamics* (Academic, 1982).

In documenting the general circulation, the authors rely almost exclusively on their own univariate analyses of rawinsonde data. I would have liked a little more emphasis on recent studies using multivariate analyses derived from operational numerical weather-prediction models. These are more reliable, particularly in the data-sparse regions of the Southern Hemisphere. I would also have preferred less emphasis on balance requirements for their own sake (such as in the case of energetics and entropy) and more extensive discussion of climate variability.

On the whole, the book is informative and authoritative on a remarkably wide range of topics. It is a fitting tribute to Victor Starr and a unique and useful addition to the literature on climate dynamics. □

John M. Wallace is in the Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195, USA.

Autumn Books

Next week's issue (26 November) contains *Nature's* Autumn Books supplement, which will feature reviews by, among others, Michael Berry (counting, thinking and being), Ruth Hubbard (the abortion controversy), Mary Warnock (university education), S. S. Schweber (Richard Feynman and modern physics), Ernest Gellner (irrationality), John Polkinghorne (science and common sense) and Lawrence Freedman (the Cold War).

Possibility of an Arctic ozone hole in a doubled- CO_2 climate

John Austin*, Neal Butchart† & Keith P. Shine‡

* Meteorological Office, London Road, Bracknell, RG12 2SZ, UK

† Hadley Centre for Climate Prediction and Research, London Road, Bracknell, RG12 2SY, UK

‡ Department of Meteorology, University of Reading, 2 Earley Gate, Reading, RG6 2AU, UK

Increased atmospheric carbon dioxide concentrations are expected to cause cooling of the lower stratosphere. This could enhance the formation of polar stratospheric clouds, which convert potential ozone-depleting species to their active forms. In an idealized three-dimensional numerical simulation of the Northern Hemisphere winter stratosphere, doubling the CO_2 concentration leads to the formation of an Arctic ozone hole comparable to that observed over Antarctica, with nearly 100% local depletion of lower-stratospheric ozone.

THE springtime reduction of ozone over Antarctica is now well established observationally^{1–3} and the processes involved are quantitatively understood theoretically⁴. Whether similar processes could also operate over the Arctic to produce a massive reduction in the ozone abundance remains uncertain. With the continued concern about the ozone layer, however, the chemical composition of the Arctic lower stratosphere has been observed extensively^{5–7} and the theoretical behaviour of polar ozone has been analysed in numerical models^{8,9}.

Measurements made during the 1989 Airborne Arctic Stratospheric Expedition (AASE) showed large abundances of ClO in the lower stratosphere¹⁰ together with low values of NO^{11} , indicating that the Arctic polar vortex was extensively perturbed by heterogeneous chemistry and was “primed for ozone destruction”. Indeed, local ozone depletion rates in the lower stratosphere inferred from the AASE observations were in excess of 1% per day^{12,13}, approaching the values calculated for extended periods in the Antarctic¹⁴. Similarly, observations of ozone during the winter and early spring^{5,6} provided circumstantial evidence for local depletion of up to ~30% in the lower stratosphere, which would imply total column changes of up to 15%. This is consistent with satellite observations in high latitudes³, although it is considerably smaller than observed over Antarctica.

The conclusions of a more comprehensive (albeit idealized) three-dimensional modelling study⁹ provide a different picture and suggest that massive ozone destruction as seen over Antarctica is unlikely to occur over the Arctic with current conditions. Although rapid ozone depletion does occur locally in the model, consistent with observations, the downward transport associated with the Northern Hemisphere meteorological conditions significantly limits the ozone depletion on a seasonal timescale. Furthermore, this downward transport is likely to limit Arctic ozone destruction despite the increases in stratospheric chlorine amounts anticipated from estimates of chlorofluorocarbon release^{8,15}. Consequently, it seems that the appearance of an Arctic ozone hole would require a considerable change in the stratospheric circulation. Increases in CO_2 anticipated over the next 50 years¹⁶ are expected to produce such a change by cooling the lower stratosphere^{17,18}. This is likely to increase the occurrences of polar stratospheric clouds (PSCs) which are an essential component for springtime polar ozone destruction⁴. Moreover, the reductions in ozone itself could produce some positive feedback by further reducing the radiative heating and hence the temperature of the lower stratosphere¹⁹.

Here we analyse the effects of increases in CO_2 and radiative feedbacks on the evolution of polar ozone in a three-dimensional model. With current concentrations of CO_2 the model provided

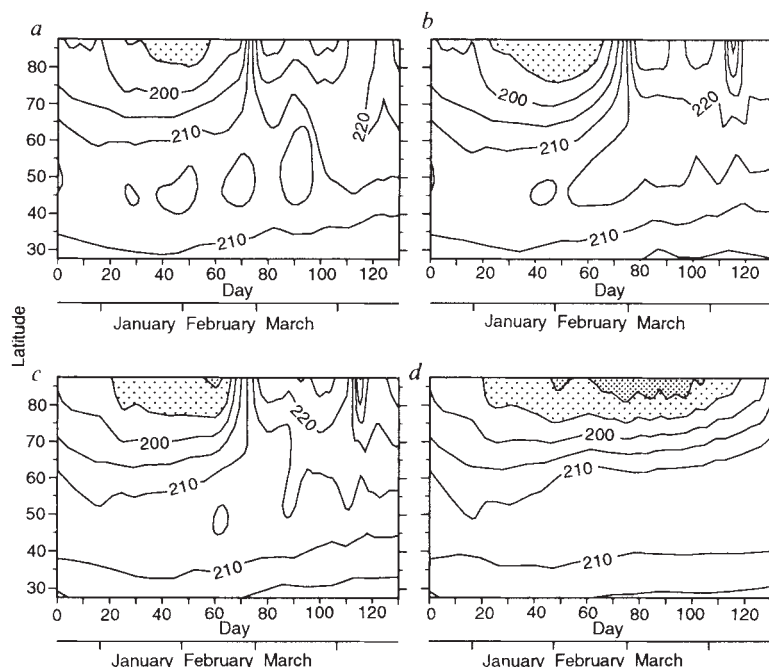


FIG. 1 Zonal mean temperature at 56 mbar for the model simulations in which the radiative heating and cooling were calculated using *a*, present-day levels of CO_2 and climatological ozone concentrations; *b*, present-day levels of CO_2 but with model-calculated ozone concentrations; *c*, doubled CO_2 with climatological ozone concentrations; and *d*, doubled CO_2 and model-determined ozone concentrations. Light shading indicates where temperatures are below 195 K and dark shading where temperatures are below 190 K.

an idealized simulation of the Northern Hemisphere winter from mid-December through to April with a relatively undisturbed circulation in January and February followed by a late stratospheric warming. This is typical of those years in which prolonged and more extensive PSC formation would be most likely, and the column ozone was reduced by ~ 40 Dobson Units (DU) where PSCs formed. When the concentration of CO_2 was doubled, the area of depleted ozone increased, but otherwise there was little change in the ozone behaviour, except when the effects of ozone radiative feedbacks were included in the model. In this situation the circulation was altered so that a cold polar vortex remained until the end of the simulation, thereby maintaining the presence of PSCs for a much longer period. This resulted in a substantial ozone change, with the column ozone reduced over a small area in mid-April to under 200 DU, approaching the low values observed in recent years over Antarctica.

Numerical experiments

We use a version of the numerical model developed in the UK Meteorological Office to describe the meteorology and photochemistry of the stratosphere and lower mesosphere⁹. We have

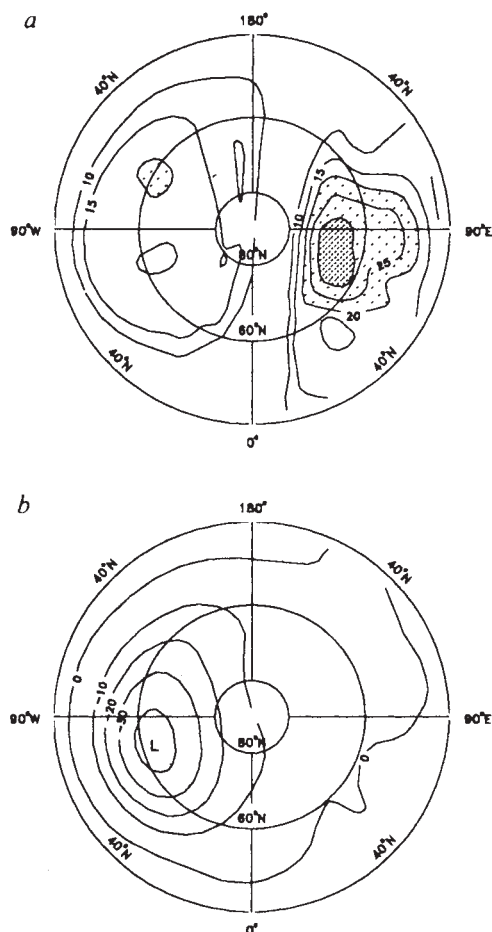


FIG. 2 *a*, Percentage change in ozone on 25 March at 56 mbar resulting solely from chemical processes acting from the start of the simulation in which the radiative heating and cooling were calculated using present-day levels of CO_2 and climatological ozone concentrations. Contours are shown only where the chemistry has depleted ozone by more than 10%. Values greater than 20% are lightly shaded and those greater than 30% more heavily shaded. *b*, Change in the total ozone (Dobson Units) due to heterogeneous processes for the period 15 December to 25 March from the simulation in which the radiative heating and cooling were calculated using present-day levels of CO_2 and climatological ozone concentrations.

improved the model by including bromine chemistry and refined the treatment of PSCs to allow for the sedimentation and evaporation as described in section 5 of ref. 9. Also, a detailed treatment of radiative heating and cooling²⁰ takes account of the concentration of CO_2 and the distribution of ozone. Both climatological and model-determined ozone fields were used. The PSC parametrization is temperature-dependent and therefore is valid for both current and doubled CO_2 situations. With such complexity 21 minutes of Cray-YMP time are required per model day, and this restricted the number and length of simulations that could be done.

Initial conditions for the chemical constituents were taken from two-dimensional model results²¹ and adjusted to be consistent with recent observations⁷. Initial ozone concentrations were taken from a mid-December climatology²⁰. For consistency, the same ozone field was used in the radiative calculation but updated during the course of the simulations using either climatological²⁰ or model-calculated values. The amount of total inorganic bromine was taken to be a linear function of the amount of total inorganic chlorine constrained by high-latitude measurements²². All the simulations used the same zonally symmetric chemical initial conditions.

The meteorological behaviour of the Northern Hemisphere winter was simulated by prescribing at the model lower boundary (316 mbar) a perturbation geopotential height field. The perturbation, although qualitatively consistent with climatological wave amplitudes^{23,24}, was chosen to enable the model to simulate a Northern Hemisphere winter in which the polar stratosphere remained relatively cold throughout January and most of February with a late stratospheric warming. The perturbation had a maximum amplitude at 60°N and decreased to zero at 82.5°N and 27.5°N . Only the first two zonal harmonics were included. The amplitudes were initially zero and increased asymptotically to 160 m and 130 m at 60°N with an e-folding timescale of 3 days.

All the simulations covered the 130-day period from 15 December. For current levels of CO_2 (330 p.p.m.v.), the initial temperature fields were obtained from a 10-year zonally-averaged December climatology derived from data retrieved from the Tiros-N series of Stratospheric Sounding Units. For doubled CO_2 the model was first integrated for perpetual 15 December conditions in zonally symmetric form without chemistry. After about 90 days there was a steady systematic temperature difference between two integrations with 330 p.p.m.v. and 660 p.p.m.v. of CO_2 , similar to that predicted in climate change studies^{17,18}. This was then used to modify the December climatology to approximate a doubled CO_2 climatology. Initial wind fields were derived from the temperature fields using the hydrostatic approximation and gradient wind balance.

The temperature changes for doubled CO_2 would be different if, instead, a mixture of increased greenhouse gas concentrations (CO_2 , CH_4 , N_2O , CFCs) was considered²⁵. Calculations with the University of Reading's Fixed Dynamical Heating Model²⁶ indicate that temperature reductions in the wintertime polar lower stratosphere are reduced by about one-third if changes in these gases are included in a simulation from 1765 to 1990, instead of the 'effective CO_2 ' for the same period. Changes in greenhouse gases and in the stratospheric circulation itself will also affect the tropospheric circulation²⁷. Nevertheless, because the interannual variations in the Northern Hemisphere planetary wave amplitudes are considerably larger than the expected climatic variability, the chosen tropospheric forcing should remain realistic over the range of climate scenarios considered here. The same tropospheric forcing was therefore applied at the model lower boundary in all the simulations.

Zonal mean temperature evolution

Figure 1 shows the evolution of the zonal mean temperature in the lower stratosphere (56 mbar) for four simulations. With

current CO_2 and with radiative heating rates calculated using climatological ozone, high-latitude mean temperatures remained low until the occurrence of a stratospheric warming in late February (Fig. 1a). This behaviour is similar to the observed evolution of the mean temperature in the lower stratosphere during 1984 and 1989. When the model-calculated ozone was used instead of the climatological values (Fig. 1b) there was little change. Similarly, there was little difference in the overall behaviour when CO_2 was doubled, with climatological ozone (Fig. 1c) but, as might be expected, the temperatures were about 3 K lower. Finally, with both doubled CO_2 and fully interactive ozone (Fig. 1d) there was a marked change in the evolution and a stratospheric warming no longer occurred. Instead mean temperatures in polar latitudes remained under 190 K throughout March, low enough for the formation of PSCs at this level. This is lower than the daily minimum temperature observed in the polar vortex in March for the years 1964–89 (ref. 28). The additional diabatic cooling occurred predominantly through a decrease in solar absorption by ozone^{19,29}.

Present-day Arctic ozone depletion

Figure 2a shows where chemical processes have reduced ozone by more than 10% in the simulation with current CO_2 but without radiatively interactive ozone. By 25 March (100 days into the simulation) ozone has been reduced in two main regions. In the western hemisphere, there is a broad region where chemical destruction exceeds 15% which can be identified with air advected from high latitudes. In the eastern hemisphere, local ozone destruction has exceeded 30% and can be identified with air advected from low latitudes. A repeat of the simulation but without the heterogeneous reactions confirmed that the western hemisphere depletion was almost entirely due to heterogeneous

processes whereas the change in the eastern hemisphere was due to homogeneous chemistry. Figure 2b shows the differences in total ozone between the two simulations on 25 March. Because there is no ozone radiative feedback in either simulation, the dynamical evolutions are identical, and therefore differences are due entirely to the changes in chemistry. The figure shows a maximum decrease of just over 40 DU at about 65° N. Similar behaviour was also obtained with fully interactive ozone, but because of the radiative coupling of the ozone field to the dynamical evolution it was not possible to identify unambiguously the contributions of heterogeneous and homogeneous chemistry. Nonetheless, the model results suggest that under current conditions heterogeneous processes involving PSCs reduce total ozone in the high-latitude Northern Hemisphere by about 40 DU and give rise to a local loss of over 20%. These results are broadly consistent with observed ozone changes^{3,5}.

PSCs and polar ozone chemistry

Figure 3a shows the fractional coverage of PSCs polewards of 25° N at 56 mbar for current CO_2 with both the noninteractive (run 1) and interactive (run 2) ozone, and for the same simulations with doubled CO_2 (runs 3 and 4). PSCs were initially absent from all the simulations but began to form first with doubled CO_2 just before day 20 (4 January) and about 10 days later with current CO_2 . For three of the simulations the area of PSCs reached a maximum at about day 50 (3 February) before reducing rapidly to zero by early March, corresponding to the rapid increase in temperatures in Fig. 1. For the remaining simulation with doubled CO_2 and radiatively interactive ozone the PSCs existed well into April.

PSCs affect ozone by perturbing the nitrogen and chlorine species⁴. Figure 3b and c shows for the two simulations with radiatively interactive ozone, the evolution of ozone, total odd nitrogen N_y and ClO at 56 mbar at the position of the temperature minimum which lies within the PSC region. With 330 p.p.m.v. of CO_2 (Fig. 3b), ozone gradually increased at 56 mbar because of downward transport, but a substantial drop in N_y occurred shortly after day 40 (24 January). The N_y recovered soon after day 60 (13 February) returning roughly to its initial value. The low N_y or denitrification occurred during the presence of PSCs but the existence of PSCs did not automatically result in denitrification because the temperatures were not always low enough (Fig. 3a and b). The concentrations of ClO remained extremely low during these periods because air at the temperature minimum was in darkness at the model output time (00:00 GMT). With doubled CO_2 (Fig. 3c) denitrification again occurred but concentrations of N_y lower than about 1 p.p.b.v. were maintained until the last 10 days of the simulation when downward transport restored the higher values. In fact at the end of March in this simulation there was denitrification at all levels from 75 to 20 mbar, similar to that observed over Antarctica. From about day 80 (5 March) ClO increased to over 0.7 p.p.b.v. in Fig. 3c, corresponding to over 1 p.p.b.v. at local noon. By this time solar illumination was sufficient even in high latitudes to drive the same catalytic ozone destruction cycle involving the ClO dimer³⁰ that is responsible for the Antarctic ozone hole⁴. As a result essentially all the ozone had been destroyed at 56 mbar at this location.

Reduction of the ozone column

Figure 3d shows the evolution of the average ozone column polewards of 75° N for the two simulations with interactive ozone (runs 2 and 4). The other two simulations were barely distinguishable from run 2 where there is a gradual increase from the initial value of 370 DU to about 500 DU on day 80. Thereafter, the mean ozone remained broadly constant. In contrast, with doubled CO_2 and fully interactive ozone (run 4) the mean column amount decreased steadily after day 80 and by the end of the simulation was ~150 DU lower than in the other simulations. Although PSCs were present early in all the

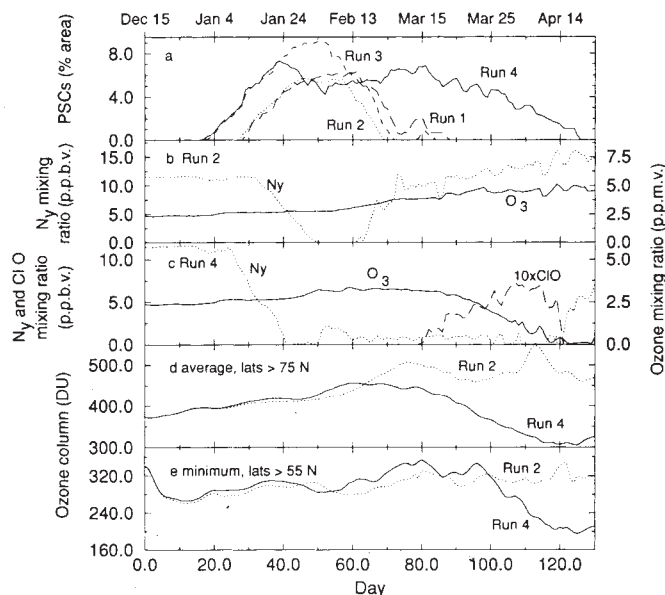
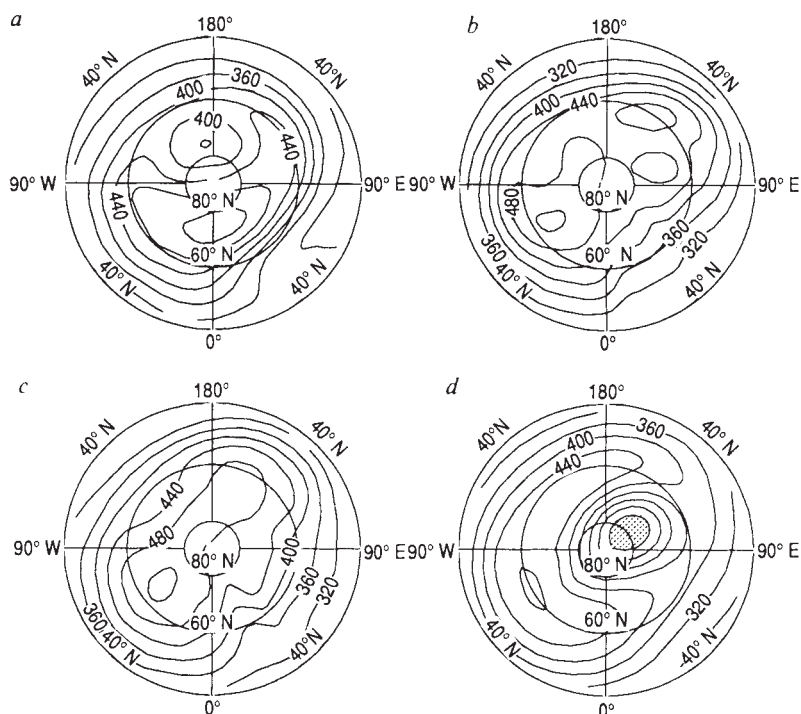


FIG. 3 a, Percentage of the model domain north of 25° N covered by polar stratospheric clouds at 56 mbar for the simulations in which the radiative heating and cooling were calculated using present-day levels of CO_2 and climatological ozone concentrations (run 1), present-day levels of CO_2 but with model-calculated ozone concentrations (run 2), doubled CO_2 with climatological ozone concentrations (run 3), and doubled CO_2 and model determined ozone concentrations (run 4). b, Concentrations of O_3 , N_y and ClO at 00:00 GMT and 56 mbar at the position of the local temperature minimum, for run 2. c, As Fig. 3b but for run 4. In b and c, the local times of each point were within a few hours of local noon and so the ClO values may be taken as qualitatively representative of the day maximum, which in run 4 reached 1.1 p.p.b.v. at this level. d, Average ozone column (Dobson Units) polewards of 75° N, and e, minimum ozone column (Dobson Units) polewards of 55° N for runs 2 and 4.

FIG. 4 Total ozone (Dobson Units, including an estimated tropospheric component) on 14 April for the corresponding model simulations *a-d* shown in Fig. 1. Shading denotes where the ozone column is under 240 DU. Because of the idealized nature of the numerical experiments the longitudinal orientation in these maps is arbitrary and is merely determined by the specified phases of the planetary waves prescribed at the model lower boundary (similarly for Fig. 2). The coastline shown on the colour picture (front cover) should therefore only be used as a guide to the size of the ozone hole and not its longitudinal position.



simulations (Fig. 3*a*), in three of these they had disappeared before having a significant effect on the ozone amounts. In the final simulation however, the continued presence of PSCs resulted in elevated levels of ClO (Fig. 3*c*) and substantial ozone destruction. Figure 3*e* shows the evolution of the minimum ozone column polewards of 55° N. Again, the behaviour in three of the simulations is very similar (only run 2 is shown) whereas with doubled CO₂ and fully interactive ozone (run 4) an ozone minimum of under 200 DU occurred in mid-April. Although not as low as the extreme values observed in some Antarctic ozone holes, it is typical of the minimum which appeared in the middle 1980s¹.

Maps of the column ozone amounts on day 120 (14 April) of each simulation are presented in Fig. 4. Again the total ozone amounts shown in Fig. 4*a-c* are very similar with maxima near 70° N varying between 526 and 541 DU. In these three figures there are no clearly defined minima in high latitudes, but the lowest values polewards of 55° N are very similar and vary between 307 and 352 DU. These low values are due mostly to horizontal transport of lower ozone amounts from mid-latitudes rather than chemical destruction. In contrast with doubled CO₂ and fully interactive ozone (Fig. 4*d*), there is a clearly defined minimum in high latitudes whilst the maximum ozone amount is about 10% lower than in the other simulations. The reduction in the mid-latitude ozone maximum has also been noted for the Antarctic³¹ and the overall figure is reminiscent of the Southern Hemisphere spring total ozone distribution³¹, albeit with an ozone hole of smaller total area.

Prospects and uncertainties

Increased concern over the atmospheric ozone layer followed from the discovery of massive springtime reductions in ozone over Antarctica¹. The effects of these depletions on the global ozone budget³² and the possibility of similar massive ozone reductions occurring over the Arctic are of prime interest. Although idealized, our three-dimensional model results provide direct evidence of a possible future Arctic ozone hole, but this will require current levels of chlorine combined with a doubling in the concentration of CO₂ to 660 p.p.m.v. With 330 p.p.m.v. of CO₂, our model results confirm earlier findings^{8,9} that an Arctic ozone hole is unlikely at present. From our necessarily

limited number of simulations, it was not possible to establish the threshold concentration of CO₂ at which the probability of an ozone hole increases. During the next few decades chlorine is expected to increase then decrease¹⁵, while CO₂ is expected to increase continuously¹⁶. It is therefore likely that a combination of conditions will occur during the next 50 years which will give rise to an Arctic ozone hole unless both CO₂ and chlorine emissions are curbed.

In the model the Arctic ozone hole appeared as a direct result of the changes in circulation induced by both doubling the amount of CO₂ and properly representing the radiative feedbacks of the evolving ozone distribution. Neither change was by itself sufficient to lead to the formation of an ozone hole. Also, because of the interannual variability of the stratosphere, it is not possible to determine accurately from the results of this study the frequency of occurrence of Arctic ozone holes. Recent calculations with doubled CO₂ using a low-resolution general circulation model³³ failed to produce an Arctic ozone hole partly, it appears, because the modelled January zonal mean temperatures in the Arctic lower stratosphere with current CO₂ were much higher than climatological. In contrast, our simulations were designed for cold winters and indeed the zonal mean temperatures in the lower stratosphere for two of the past 10 years (1984, 1989) were similar to the 330-p.p.m.v. results presented here. Therefore it might be expected that up to 20% of winters will produce Arctic ozone holes for doubled CO₂ conditions. For the remaining 80% of winters, it is unlikely that doubling of CO₂ will suppress the occurrence of a stratospheric warming sufficiently to produce an Arctic ozone hole, although even in these conditions ozone depletion could be enhanced³³.

Finally, there are other potentially important ways in which the polar ozone destruction mechanism may be perturbed in future but which have not been considered here. Changes in the tropospheric circulation will affect planetary wave propagation and hence the temperature variability of the lower stratosphere. Changes in stratospheric water vapour due to changes in methane³⁴ and stratosphere-troposphere exchange could affect both the radiative budget and the temperature of PSC formation. A long-term decrease in ozone might systematically affect stratospheric temperatures so that the beginning of winter could be colder than in our simulations. Nevertheless, the occur-

rence of an Arctic ozone hole directly as a result of increased CO₂ could be one of the more serious consequences of climate change. Results presented here together with observations of lower stratospheric temperatures suggest that the interannual variability of ozone depletion in the Northern Hemisphere

spring will increase considerably in the future and is likely to be determined largely by the meteorological behaviour of the upper troposphere. Therefore, in considering the ramifications of climate change, it will be important to predict details of the upper air circulation as well as mean surface parameters. □

Received 15 May; accepted 15 October 1992.

- Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207–210 (1985).
- Gardiner, B. G. *Geophys. Res. Lett.* **15**, 901–904 (1988).
- Stolarski, R. S., Bloomfield, P., McPeters, R. D. & Herman, J. R. *Geophys. Res. Lett.* **18**, 1015–1018 (1991).
- Solomon, S. *Nature* **347**, 347–354 (1990).
- Koike, M. *et al.* *Geophys. Res. Lett.* **18**, 791–794 (1991).
- Proffitt, M. H. *et al.* *Nature* **347**, 31–36 (1990).
- Geophys. Res. Lett.* **17**, No. 4 (1990).
- Granier, C. & Brasseur, G. *J. geophys. Res.* **96**, 2995–3011 (1991).
- Austin, J. & Butchart, N. *J. geophys. Res.* **97**, 10165–10186 (1992).
- Brune, W. H., Toohey, D. W., Anderson, J. G. & Chan, K. R. *Geophys. Res. Lett.* **17**, 505–508 (1990).
- Fahey, D. W., Kawa, S. R. & Chan, K. R. *Geophys. Res. Lett.* **17**, 489–492 (1990).
- Jones, R. L., McKenna, D. S., Poole, L. R. & Solomon, S. *Geophys. Res. Lett.* **17**, 549–552 (1990).
- McKenna, D. S. *et al.* *Geophys. Res. Lett.* **17**, 553–556 (1990).
- Austin, J. *et al.* *J. geophys. Res.* **94**, 16717–16735 (1989).
- SORC *Stratospheric Ozone 1991* (HMSO, London, 1991).
- Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (eds) *Climate Change, the IPCC Scientific Assessment* (Cambridge Univ. Press, Cambridge, 1990).
- Feis, S. B., Mahlman, J. D., Schwarzkopf, M. D. & Sinclair, R. W. *J. atmos. Sci.* **37**, 2265–2297 (1980).
- Rind, D., Suozzo, R., Balachandran, N. K. & Prather, M. J. *J. atmos. Sci.* **47**, 475–494 (1990).

- Shine, K. P. *Geophys. Res. Lett.* **13**, 1331–1334 (1986).
- Shine, K. P. *Q. J. R. met. Soc.* **115**, 265–292 (1989).
- Garcia, R. R. & Solomon, S. *J. geophys. Res.* **88**, 1379–1400 (1983).
- Toohey, D. W., Anderson, J. G., Brune, W. H. & Chan, K. R. *Geophys. Res. Lett.* **17**, 513–516 (1990).
- Geller, M. A., Wu, M.-F. & Gelman, M. E. *J. atmos. Sci.* **41**, 1726–1744 (1984).
- Randel, W. J. *Q. J. R. met. Soc.* **114**, 1385–1409 (1988).
- Wang, W.-C., Dudek, M. P., Liang, X.-Z. & Kiehl, J. T. *Nature* **350**, 573–577 (1991).
- Ramaswamy, V., Schwarzkopf, M. D. & Shine, K. P. *Nature* **355**, 810–812 (1992).
- Boville, B. A. *J. atmos. Sci.* **41**, 1132–1142 (1984).
- Nagatani, R. M., Miller, A. J., Gelman, M. E. & Newman, P. A. *Geophys. Res. Lett.* **17**, 333–336 (1990).
- Kiehl, J. T., Boville, B. A. & Briegleb, B. P. *Nature* **332**, 501–504 (1988).
- Molina, L. T. & Molina, M. J. *J. phys. Chem.* **91**, 433–436 (1987).
- Schoeberl, M. R., Stolarski, R. S. & Krueger, A. J. *Geophys. Res. Lett.* **16**, 377–380 (1989).
- Prather, M., Garcia, M. M., Suozzo, R. & Rind, D. *J. geophys. Res.* **95**, 3449–3471 (1990).
- Pitari, G., Palmeri, S., Visconti, G. & Prinn, R. G. *J. geophys. Res.* **97**, 5953–5962 (1992).
- Jones, R. L. *et al.* *Q. J. R. met. Soc.* **112**, 1127–1143 (1986).

ACKNOWLEDGEMENTS. We thank M. D. Schwarzkopf (Geophysical Fluid Dynamics Laboratory, NOAA) for updated CO₂ transmittance matrices and J. D. Mahlman for comments. The Hadley Centre is a research unit of the Meteorological Office.

Mutations in T-cell antigen receptor genes α and β block thymocyte development at different stages

Peter Mombaerts*, Alan R. Clarke†, Michael A. Rudnicki‡, John Iacomini*, Shigeyoshi Itohara*, Juan J. Lafaille*, Lili Wang*, Yoshiaki Ichikawa*, Rudolf Jaenisch‡§, Martin L. Hooper†|| & Susumu Tonegawa*¶

* Howard Hughes Medical Institute at the Center for Cancer Research, and § Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Cancer Research Campaign Laboratories, Department of Pathology, and || AFRC Centre for Genome Research, Edinburgh University, Edinburgh EH8 9AG, UK

‡ Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

Analysis of mice carrying mutant T-cell antigen receptor (TCR) genes indicates that TCR- β gene rearrangement or expression is critical for the differentiation of CD4⁺CD8⁺ thymocytes to CD4⁺CD8⁺ thymocytes, as well as for the expansion of the pool of CD4⁺CD8⁺ cells. TCR- α is irrelevant in these developmental processes. The development of $\gamma\delta$ T cells does not depend on either TCR- α or TCR- β .

IN the vertebrate immune system the recognition of a large variety of antigens is achieved by antigen-specific lymphocyte receptors, the immunoglobulin on B lymphocytes¹ and the T-cell antigen receptor on T lymphocytes². There are two types of TCR, $\alpha\beta$ TCR and $\gamma\delta$ TCR, which are expressed on the surface of distinct subsets of T lymphocytes³ that seem to be derived from different cell lineages⁴.

In the main pathway of $\alpha\beta$ T-cell development in the adult thymus, bone marrow-derived stem cells undergo a series of intermediate stages. Among these stages are CD4⁺CD8⁺IL-2R⁺TCR⁺ (double negative, or DN) and CD4⁺CD8⁺IL-2R⁺TCR^{low+} (double positive, or DP) cells, both of which are important control points at which some signal or stimulus from the thymic microenvironment is suspected or known to be required for proliferation or differentiation to proceed⁵. The end stage of the pathway is represented by CD4⁺CD8⁺TCR^{high+} or

CD4⁺CD8⁺TCR^{high+} cells (both are said to be in single positive, or SP, stages), which are mostly immediate precursors of major histocompatibility complex (MHC) class II-restricted helper T cells and MHC class I-restricted cytotoxic T cells, respectively.

It is well established that $\alpha\beta$ TCR play critical roles in the DP to SP transition by interacting with self-ligands expressed on the surface of thymic stroma cells (positive selection). But the role of the TCR genes or their products in the earlier DN to DP transition is uncertain. Because in the developing fetal thymus TCR- β genes rearrange before TCR- α genes^{6,7}, a distinct role can be envisaged for each of these genes or their products. Indeed, analysis of *severe combined immunodeficiency* (*scid*) mice carrying rearranged TCR transgenes indicated that TCR- β gene rearrangement or expression could be instrumental in the DN to DP transition^{8,9}. But it is not known whether TCR- β gene rearrangement or expression is required for such a transition, and no information is yet available on a role for TCR- α gene rearrangement or expression.

¶ To whom correspondence should be addressed.

Here we use gene targeting to introduce a mutation in the TCR α or β genes into the germ line of mice. Initial analysis of these mice revealed that TCR- β gene rearrangement or expression is necessary for the maintenance of normal thymocyte number and the DN to DP transition. Our results show that a single rearranged TCR- β gene introduced as a transgene can quantitatively convert DN cells to DP cells from DN cells that have accumulated in mice mutant for the recombination activating gene-1 (*RAG-1*)¹⁰ and increase the total thymocyte number to the normal level. By contrast, TCR- α gene expression seems to be irrelevant for these developmental events. Our study also indicates that $\alpha\beta$ T cells are unnecessary for the generation of apparently normal $\gamma\delta$ T cells.

Creation of mutations in TCR genes

The creation of a deletion at the TCR- β locus in embryonic stem (ES) cells has been described¹¹. Briefly, two ES clones were produced carrying a genomic deletion of 15 kilobases (kb) that encompasses the TCR- β locus from $\Delta\beta 1.3$ to downstream of *C β 2*. Germ-line chimaeras were obtained from both clones. A line was established from clone 59 and clean stocks derived by caesarean section.

To produce a targeted disruption of the unique TCR α segment, we constructed targeting vector pPMKO-1 using 3.9 kb of homologous DNA sequences of BALB/c origin and carrying the *pgk-neo* selectable marker in the first exon of TCR α (Fig. 1a). Two out of 160 G418-resistant D3 clones¹² were found to have the targeted mutation and clone 515 gave rise to germ-line chimaeras. For both the TCR- α and TCR- β mutations, intercrosses between heterozygous mice gave rise to homozygous mutant mice at the expected frequency of 25%. An example of the results from Southern analysis is shown in Fig. 1b for both TCR- α and TCR- β littermates.

Numbers of thymocytes

The numbers of total thymocytes were determined for TCR- α and TCR- β mutant mice, and for mice doubly homozygous for the TCR- α and TCR- β mutations (called TCR $\alpha \times \beta$ mutant mice). Whereas the thymuses of TCR- α mutant mice contained similar numbers of cells to those of normal littermates (Fig. 2a), the thymuses of the TCR- β mutant mice contained fewer cells, ranging from 6- to 60-fold less than those of normal littermates (Fig. 2a). TCR $\alpha \times \beta$ mutant mice had similar numbers of thymocytes as TCR- β mutant mice (Fig. 2a).

The importance of TCR- β in regulating the total number of thymocytes was also demonstrated by crossing a functionally rearranged TCR- β transgene^{13,14} into the *RAG-1* mutant mice. This resulted in restoration of the numbers of total thymocytes to wild-type levels (Fig. 2a). These results indicate that TCR- β is necessary to generate a thymus with the wild-type number of cells, and that it alone can lift the blockade imposed by the *RAG-1* mutation in this respect. In contrast, TCR- α is irrelevant to the generation of wild-type numbers of thymocytes.

Blockade by TCR- α mutation

As shown in Fig. 3a, thymuses of TCR- α mutant mice are largely devoid of CD4 or CD8 SP cells but retain normal numbers of small DP cells without the interleukin-2 (IL-2) receptor (IL-2R⁺ DP cells) (Fig. 2b). The latter cells do not express $\alpha\beta$ TCR, or only barely, as determined by staining with the pan-TCR- β monoclonal antibody H57-597 (ref. 15) (Fig. 3a). There is also no difference in the number of DN cells in TCR- α mutant mice as compared to wild-type littermates (Fig. 3a).

The overall results indicate that expression of a rearranged TCR- α chain is unnecessary either for the progression of thymocytes from the DN to DP stage or for the expansion of the

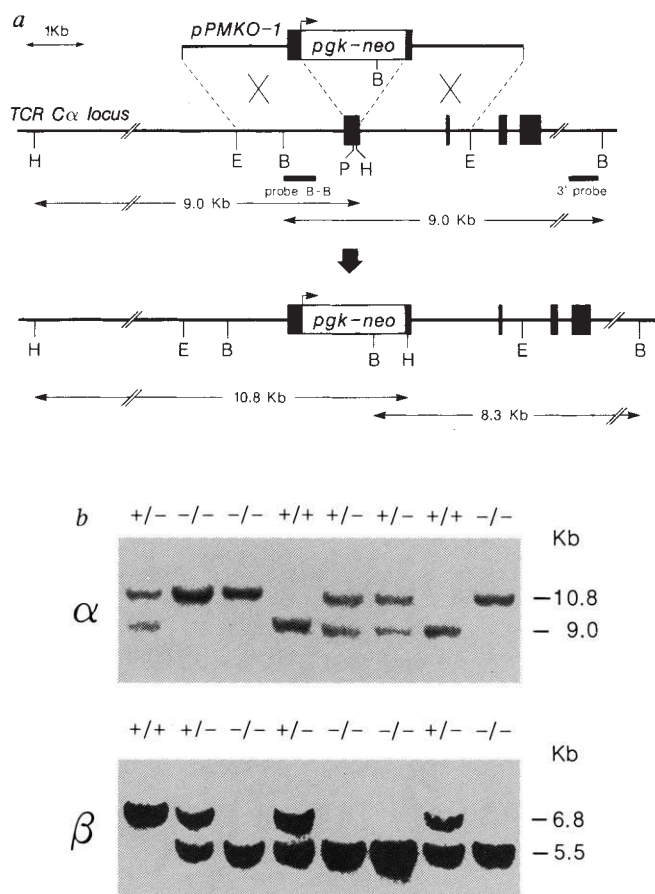


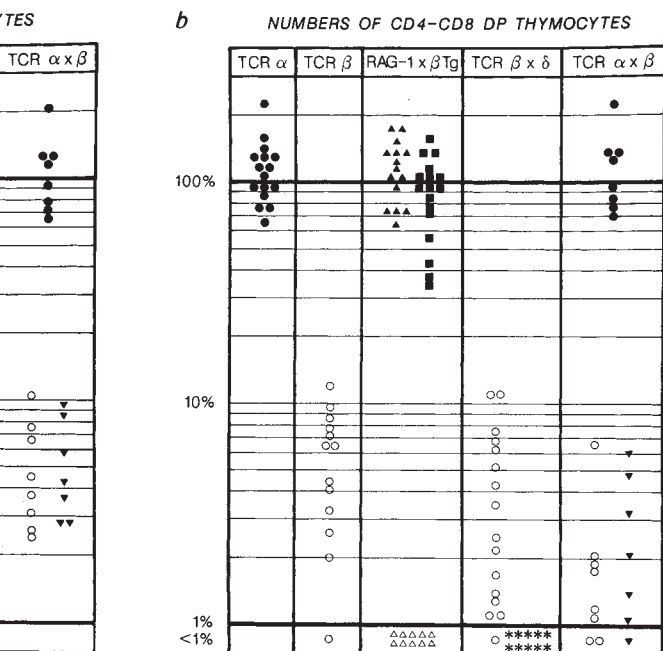
FIG. 1 Targeting of TCR α and Southern analysis of TCR- α and TCR- β mutant mice. *a*, *top*, TCR- α targeting construct pPMKO-1. Crossed lines indicate hypothetical crossovers between the targeting construct and the TCR α locus, *middle*, genomic structure of the TCR α locus. Exon I codes for most of the extracellular domain, exon II encodes the hinge domain, exon III encodes the transmembrane and cytoplasmic domains, and exon IV contains untranslated sequences; *bottom*, structure of the disrupted TCR α locus. Upon targeted integration of pPMKO-1, the *pgk-neo* gene is inserted in the first exon of TCR α . Probes are probe B-B, to genotype the mice, and 3' β probe, to screen the ES clones. Sizes of DNA fragments hybridizing to these probes are indicated within double-headed lines. At both left and right ends, the scale is interrupted. Abbreviations for restriction enzyme sites: H, *HindIII*; E, *EcoRI*; B, *BamHI*; P, *PfI*. *b*, Southern blots. *Top*, TCR- α mice. Thymocyte DNA from eight littermates was cut with *HindIII* and hybridized with probe B-B. The wild-type allele is at 9.0 kb and the mutant allele is at 10.8 kb. *Bottom*, TCR- β mice. Tail DNA from eight littermates was cut with *PstI* and hybridized with a 3' *C β* probe. The wild-type allele is at 6.8 kb and the mutant allele is at 5.5 kb; +/+, wild-type; +/-, heterozygous; and -/-, homozygous mutant.

METHODS. pPMKO-1 was constructed from phage λ 3.9 of BALB/c origin²³. The 3.9-kb *EcoRI* insert of this phage was subcloned into pUC18. The *pgk-neo* selectable marker²⁴ was excised as a *EcoRI-HindIII* fragment from pKJ1²⁵, blunt-ended with Klenow DNA polymerase and inserted into the unique *PfI* site in the first exon of TCR α , which was blunted by treatment with T4 DNA polymerase. Before electroporation that targeting construct was released from plasmid sequences by digestion with *EcoRI*. The targeting experiment was done as described¹⁰. ES colonies were screened by digesting with *BamHI* and hybridizing with the 3' probe, a 0.5-kb *PvuII* fragment containing the TCR- α enhancer²⁶. Chimaeric mice were generated by standard protocols²⁷. DNA was isolated according to ref. 28.

FIG. 2 Thymocyte numbers. *a*, Numbers of total thymocytes. As there is much variation in numbers of thymocytes between mice of different litters of the same age but much less between mice belonging to the same litter, the numbers of total thymocytes are shown in comparison to wild-type or heterozygous littermates in a logarithmic scale. TCR- α , TCR- α mutant mice ($\bullet$); $n=16$, average 101%. TCR- β , TCR- β mutant mice ($\circ$); $n=13$, average 8.3%. RAG-1 $\times$ β Tg: RAG-1 mutant mice (Δ), $n=10$, average 1.4%; TCR- β transgenic mice ($\blacksquare$), $n=17$, average 92%; TCR- β transgenic RAG-1 mutant mice ($\blacktriangle$), $n=15$, average 102%. TCR $\beta \times \delta$: TCR- β mutant mice, $n=16$, average 6.6%; TCR $\beta \times \delta$ mutant mice ($*$), $n=10$, average 3.4%. TCR $\alpha \times \beta$: TCR- β mutant mice, $n=8$, average 5.2%; TCR- α mutant mice, $n=8$, average 110%; TCR $\alpha \times \beta$ mutant mice (∇), $n=7$, average 5.3%.

b, Numbers of CD4-CD8 double-positive thymocytes. Data are shown for the same mice as in *a*. Percentage of DP thymocytes was measured by FACScan. TCR- α : $n=16$, average 112%; TCR- β : $n=13$, average 5.8%; RAG-1 $\times$ β Tg: RAG-1 mutant mice, $n=10$, average 0%; TCR- β transgenic mice, $n=17$, average 88%; TCR- β transgenic RAG-1 mutant mice, $n=15$, average 115%; TCR $\beta \times \delta$: TCR- β mutant mice, $n=16$, average 4.2%; TCR $\beta \times \delta$ mutant mice, $n=10$, average 0.04%. TCR $\alpha \times \beta$: TCR- α mutant mice, $n=8$, average 118%; TCR- β mutant mice, $n=8$, average 2%; TCR $\alpha \times \beta$ mutant mice, $n=7$, average 2.7%.

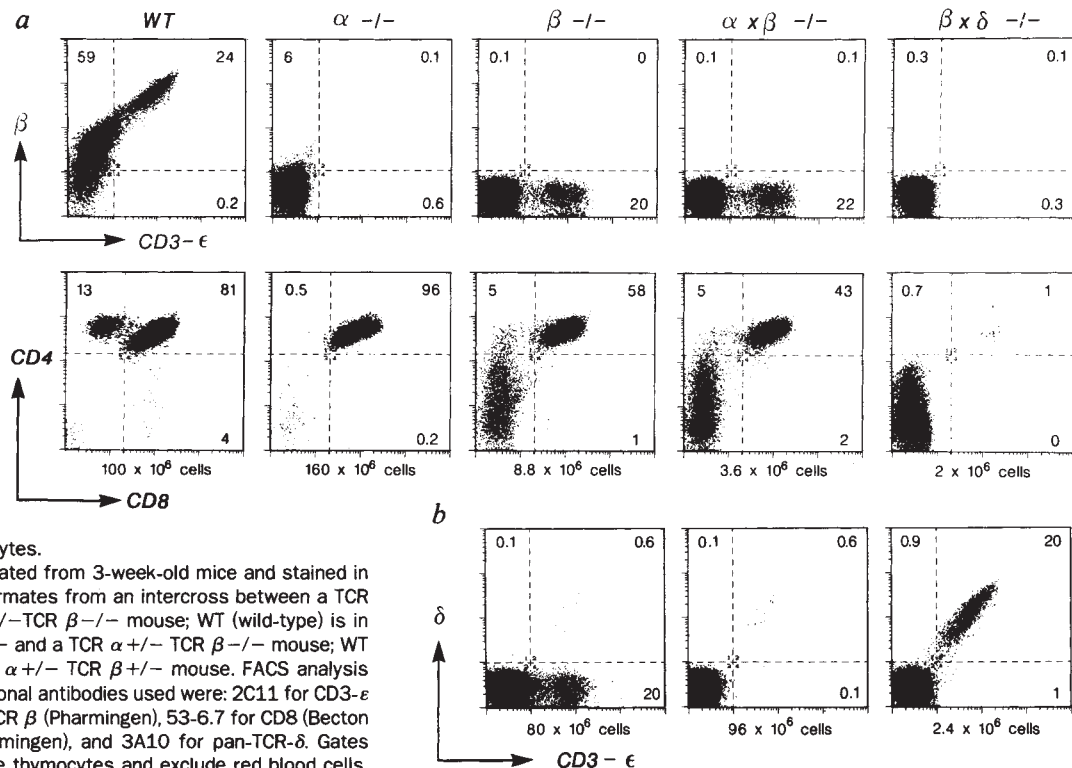
METHODS. Mice were between three weeks and three months old. The background is (129/Sv $\times$ C57BL/6) for TCR- α and RAG-1 mutant mice, and (129/Ola $\times$ Balb/c) for TCR- β and TCR- δ mutant mice. We have also obtained TCR- β and TCR- δ mutant mice in a (129/Ola $\times$ C57BL/6) background and have found no effect of genetic background on thymocyte numbers and percentage of DP cells (data not shown). TCR- β transgenic mice were of



founder 101 (ref. 14). The non-erythroid cells isolated from the thymuses were counted using a haemocytometer. At least one and usually two or more wild-type or heterozygous littermates were analysed in parallel with the mutant mice. The average number of total thymocytes of these littermates was given the value of 100%, and the number of total thymocytes in individual mutant mice was converted into a percentage. The percentage of DP thymocyte was measured by FACScan analysis by staining with 53-6.7 (Becton Dickinson) for CD8 and GK1.5 (Pharmingen) for CD4. The number of DP thymocytes was calculated by multiplying the number of total thymocytes with the percentage of DP thymocytes. This number was converted into a percentage using the wild-type or heterozygous littermates as 100%. In the case of TCR $\beta \times \delta$ mutant mice, TCR- δ mutant mice were included in the reference group of wild-type or heterozygous littermates, as they have normal numbers of total thymocytes and of DP thymocytes (our unpublished observations).

FIG. 3 FACScan analysis of thymocytes of mutant mice. *a*, top, CD3- ϵ -fluorescein isothiocyanate (FITC) (horizontally) and TCR- β -phycoerythrin (PE) (vertically); bottom, CD8-FITC (horizontally) and CD4-biotin-PE (vertically). WT, Wild-type; $\alpha^{-/-}$, TCR- α mutant; $\beta^{-/-}$, TCR- β mutant; $\alpha \times \beta^{-/-}$, TCR $\alpha \times \beta$ mutant; $\beta \times \delta^{-/-}$, TCR $\beta \times \delta$ mutant mouse. Percentage of cells in quadrants 1, 2 and 4 are indicated in the corners. The DN TCR-negative thymocytes in the TCR- β mutant mice are predominantly IL-2 receptor- and Sca-1 positive, and class I MHC high (data not shown). *b*, CD3- ϵ -FITC (horizontally) and TCR- δ -biotin-PE (vertically). Indicated below are the absolute numbers of thymocytes.

METHODS. Thymocytes were isolated from 3-week-old mice and stained in parallel. The mice in *b* were littermates from an intercross between a TCR $\alpha^{-/-}$ -TCR $\beta^{+/-}$ and a TCR $\alpha^{+/-}$ -TCR $\beta^{-/-}$ mouse; WT (wild-type) is in this case a TCR $\alpha^{-/-}$ -TCR $\beta^{+/-}$ and a TCR $\alpha^{+/-}$ -TCR $\beta^{-/-}$ mouse; WT (wild-type) is in this case a TCR $\alpha^{+/-}$ -TCR $\beta^{+/-}$ mouse. FACS analysis was done as described²⁰. Monoclonal antibodies used were: 2C11 for CD3- ϵ (Pharmingen), H57-597 for pan-TCR β (Pharmingen), 53-6.7 for CD8 (Becton Dickinson), GK1.5 for CD4 (Pharmingen), and 3A10 for pan-TCR- δ . Gates accommodate both small and large thymocytes and exclude red blood cells.



DP cells to the levels of wild-type mice. Thus the role of TCR- α expression seems to be confined to further differentiation to the SP stage.

Finally, as in the thymuses of wild-type littermates, about 1% of the cells bear $\gamma\delta$ TCR in the thymus of the TCR- α mutant mice, as determined by the anti-TCR- δ monoclonal antibody 3A10 (ref. 3) (Fig. 3b). These results suggest that TCR- α expression is unnecessary for the generation of $\gamma\delta$ thymocytes.

Blockade by TCR- β mutation

The thymuses of TCR- β mutant mice are different from those of wild-type littermates, not only in the total cell number (on average ~8% of wild type) but also in their cellular composition. First, as expected, no TCR- β -positive cells were detected by antibody H57-597 (Fig. 3a). Second, only about 50% of the thymocytes were double positive, although there was substantial individual variation (15–80%). This amounts to, on average, only ~6% of DP cells in the wild-type littermates (compare with Fig. 2b). As in wild-type littermates, these cells are small and negative for IL-2R (data not shown). Third, the proportion of large DN cells has increased to about 50% of the total cell number. As in wild-type littermates this thymocyte population includes two subpopulations of similar cell numbers: immature, IL-2R-positive, TCR-negative cells, and mature, IL-2R-negative, $\gamma\delta$ TCR-positive cells (data not shown). The increase in the proportion of DN cells in TCR- β mutant mice is due to the reduction in the numbers of DP cells: the number per thymus of each DN subpopulation is not altered by the introduction of the TCR- β mutation (Figs 2a, 3a and b). Finally, the TCR- β mutant thymus contains a few CD4⁺ SP and CD8⁺ SP cells, of which between half and two-thirds are $\gamma\delta$ T cells (data not shown). The rest are probably cells on the way to the DP stage⁵.

The severe reduction of DP cells (about 6% of DP cells in the wild-type littermates) by the introduction of the TCR- β mutation indicates that TCR- β rearrangement or expression is important in the DN to DP transition in the principal differentiation pathway of $\alpha\beta$ T cells. To evaluate whether TCR- β rearrangement or expression is necessary for this transition, we introduced the TCR- β mutation to TCR- δ mutant mice (to be described elsewhere), in which we found that $\alpha\beta$ T cells develop normally and normal numbers of DN, DP and SP thymocytes are present. But introduction of the TCR- β mutation into the TCR- δ mutant mice abolishes virtually all DP cells and eliminates SP cells entirely (Fig. 3a), indicating that TCR- β rearrangement or expression is required for the DN to DP transition, at least in the principal differentiation pathway of $\alpha\beta$ T cells.

We investigated whether this role of TCR- β in $\alpha\beta$ thymocyte differentiation supersedes rearrangement or expression of TCR- α gene by crossing the TCR- α and TCR- β mutant mice. As shown in Figs 2a, b and 3a, the thymuses of the TCR $\alpha \times \beta$ mutant mice were similar to those of the TCR- β mutant mice by our criteria. The mutation in TCR- β is therefore epistatic to the mutation in TCR- α .

In conclusion, the mutation in TCR- β blocks $\alpha\beta$ thymocyte differentiation at an earlier stage than the mutation in TCR- α : at the transition from the large, IL-2R-positive DN stage to the small, IL-2R-negative DP stage. As a result the total number of thymocytes is much reduced. In contrast, the differentiation of $\gamma\delta$ thymocytes does not seem to be affected by the TCR- β mutation.

TCR- β can induce DN–DP transition

As we reported previously, $\alpha\beta$ T-cell differentiation is blocked at a DN stage in RAG-1 mutant mice¹⁰. Introduction of the transgenic TCR- β gene not only restored the total number of thymocytes to the levels of wild-type mice (Fig. 2a), but allowed nearly complete transition of the large, IL-2R-positive, DN cells to the small, IL-2R-negative DP cells (Fig. 2b). These results demonstrate that a single productively rearranged TCR- β gene

provides a sufficient signal for the transition of $\alpha\beta$ lineage cells from a DN stage (at least the stage reached in the RAG-1 mutant mice) to a DP stage, and for the expansion of these DP cells to the normal level.

TCR- α rearrangement

TCR- β genes are rearranged and their RNA has been detected before TCR- α gene rearrangement in the fetal thymus^{6,7}. A signal may therefore be provided by TCR- β gene rearrangement or its RNA or protein product which is required for TCR- α gene rearrangement, but there is no direct evidence to support this hypothesis. We investigated rearrangement of TCR- α genes in the thymuses of the TCR- β mutant and other mice by Southern blot analysis using two probes simultaneously, a 5'J α 1 probe and a C α probe. Typical results are shown in Fig. 4a. In the thymuses of the wild-type and TCR- β (as well as TCR- α) heterozygous mice, rearrangement has occurred at nearly half of the TCR- α alleles (see Fig. 4 legend for definition of the rearrangement index, RI(α)). Although somewhat reduced, the rearrangement is substantial in the thymuses of TCR- β (as well as TCR- α) mutant mice. The reproducibility of these data is shown in shown in Fig. 4 legend. In support of the Southern blot data, thymuses of TCR- α mutant mice contain, albeit at a reduced level compared to the thymuses of wild-type or other mice, TCR- α transcripts of sizes expected from the fully rearranged gene (Fig. 4b). Cloning and sequencing experiments confirmed that these transcripts are indeed derived from productively or non-productively rearranged TCR- α genes (data not shown). We conclude that TCR- β rearrangement is unnecessary for TCR- α rearrangement.

TCR- β rearrangement

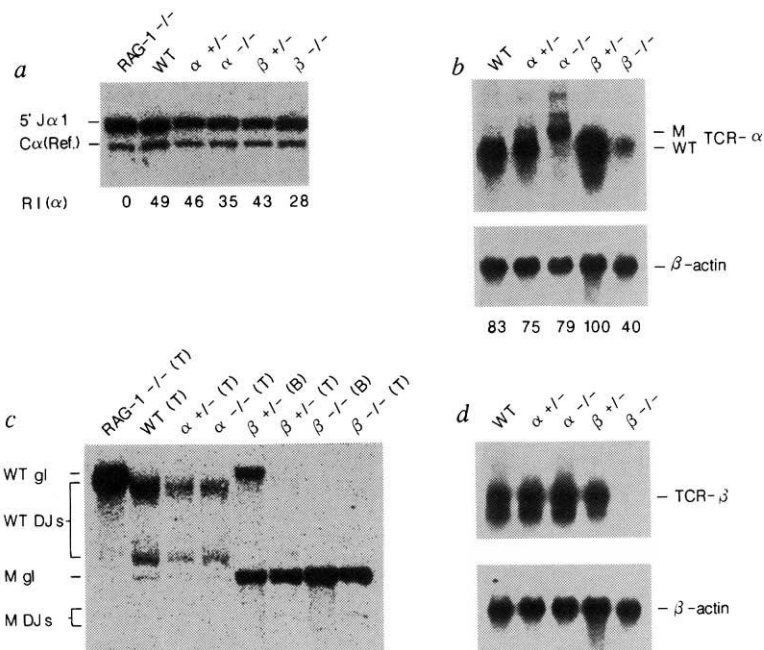
Using a 5'D β 1 probe¹⁶ we examined at TCR- β rearrangement in the different mice, including RAG-1 mutant mice in which no TCR rearrangement occurs¹⁰ (Fig. 4c). TCR- β rearrangement in the thymocytes of TCR- α mutant mice was as extensive as that in wild-type or TCR- α heterozygous mice, as indicated by the virtual disappearance of the germ-line band and the presence of several bands corresponding to DJ rearrangements. The levels of TCR- β transcripts corroborate these rearrangement results (Fig. 4d), so we confirm that TCR- α expression is not needed for TCR- β rearrangement and transcription.

In contrast to thymocyte DNA from TCR- α mutant mice, thymocyte DNA from TCR- β mutant mice gave rise to no detectable DJ bands and a mutant germ-line band as intense as brain DNA (negative control) from the same mutant mice. This finding is interesting because the mutant allele retains the intact D β 1, J β 1.1 and J β 1.2 segments despite the 15-kb deletion¹¹, and therefore can potentially undergo extensive DJ or VDJ rearrangements. The lack of rearrangement may be caused by a loss of an as-yet unidentified cis-acting rearrangement-promoting element, as suggested by the dramatic differences in the extent of rearrangement in the wild-type and mutant alleles coexisting in the same thymocytes of TCR- β heterozygous mice (Fig. 4c, third lane from right).

TCR- β surface expression

Extensive rearrangement and transcription of the TCR- β locus in TCR- α mutant mice provides a potential for synthesis of TCR- β polypeptide chains and even their cell-surface expression on thymocytes. Indeed, a small fraction of thymocytes of TCR- α mutant mice seems to be stained weakly by the anti-TCR- β antibody H57-597 (Figs 3a and 5). Surface expression of TCR- β chains was clearly seen with thymocytes derived from TCR- α mutant mice or RAG-1 mutant mice, into which a transgenic TCR- β gene was introduced; virtually all of these thymocytes were stained with H57-597 at levels higher than the TCR- β surface-positive thymocytes from TCR- α mutant mice (Figs 3a and 5). The occurrence of these cells in TCR- β transgenic RAG-1 mutant mice suggests that

FIG. 4 a, Southern blot analysis of TCR- α rearrangements. Thymocyte DNAs from 3–8 weeks old mice were cut with *SacI* and *HindIII* and were hybridized simultaneously with a 5'J α 1 probe and a TCR C α probe. Lane 1, RAG-1 mutant; lane 2, wild-type; lane 3, TCR- α heterozygous; lane 4, TCR- α homozygous mutant; lane 5, TCR- β heterozygous, and lane 6; TCR- β homozygous mutant. As all *VJ* rearrangements will delete the 5'J α 1 sequence but retain the C α sequence, the ratio of the intensity of the 5'J α 1 band to that of the C α band is inversely proportional to the extent of rearrangements in the TCR- α locus. The rearrangement index for the TCR- α locus, $RI(\alpha)$, is defined by $(1 - (5'J\alpha1/C\alpha \text{ ratio of sample}) / (5'J\alpha1/C\alpha \text{ ratio of RAG-1 sample})) \times 100$. A value of zero means no *V\alpha J\alpha* rearrangement. The $RI(\alpha)$ for the TCR- β mutant thymocyte sample is 28, meaning that 28% of the chromosomes have undergone a *V\alpha J\alpha* rearrangement. 5'J α 1/C α ratios (shown in table below) are given for different DNA samples. Three different RAG-1 mutant and seven different TCR- β mutant samples are analysed, prepared independently from pools of 3–10 mice each. The TCR- β heterozygous samples were derived from individual mice. Some samples are analysed in more than one experiment. Ratios vary between different experiments, depending on the specific activity of the two radioactively labelled probes. **b**, Northern blot analysis of TCR C α transcripts in thymocytes. *Top*, hybridization with a TCR C α probe. *Bottom*, hybridization with a β -actin probe as internal control. Lane 1, C57BL/6; lane 2, TCR- α heterozygous; lane 3, TCR- α mutant; lane 4, TCR- β heterozygous; lane 5, TCR- β mutant mice. Transcripts hybridizing to the TCR C α probe in the TCR- α mutant mouse are longer than those in the other mice, presumably because of insertion of the 1.8-kb *neo*-selectable marker in the first exon of TCR C α . WT and M designate transcripts from rearranged wild-type (1.8 kb) and mutant TCR- α alleles, respectively. Numbers at the bottom indicate the ratio of the intensity of the TCR C α band to the β -actin band, normalized to 100 for the sample of the TCR- β heterozygous thymocytes. **c**, Southern blot analysis of TCR- β rearrangements. Thymocyte (T) DNAs from 3–8-week old mice cut with *HindIII* were hybridized with a 5'D β 1 probe. Lane 1, RAG-1 mutant; lane 2, C57BL/6; lane 3, TCR- α heterozygous; lane 4, TCR- α mutant; lane 6, TCR- β heterozygous; lane 8, TCR- β mutant mice. For negative controls, brain (B) DNA was isolated from: lane 5, TCR- β heterozygous; lane 7, TCR- β mutant mice. 'WT' and 'M' mark the germ-line (unrearranged) wild-type and mutant bands, respectively. 'Gl', germ-line, and 'DJs', D-J rearrangements. The 5'D β 1 probe¹⁶ detects *DJ* rearrangements involving the D β 1 segment by the appearance of bands of known restriction fragment sizes, and *VDJ* rearrangement by a reduction in the intensity of the band representing unrearranged alleles. Rearrangement is extensive in thymocytes of wild-type compared to RAG-1 mutant mice (negative control), as indicated by the virtual disappearance of the germ-line band and the presence of several bands corresponding to *DJ* rearrangements. TCR- β rearrangement in thymocytes of TCR- α mutant mice is as extensive as that in wild-type mice. Upon prolonged



exposure, two faint bands expected from D β 1-J β 1.1 or D β 1-J β 1.2 rearrangements of the mutant allele were occasionally observed. **d**, Northern blot analysis of TCR C β 2 transcripts in thymocytes. *Top*, hybridization with TCR C β 2 probe. *Bottom*, hybridization with a β -actin probe as internal control. Same RNA samples as in **b**. TCR C β 2 transcripts were present at normal levels in TCR- α mutant mice, but were absent in TCR- β mutant mice, as the TCR C β 2 gene segment is included in the 15-kb deletion¹¹. **METHODS**. For Southern blot analysis, thymocyte and brain DNA was isolated according to ref. 28. The 5'D β 1 probe has been described before^{10,16}; the 5'J α 1 probe is a 1.2-kb *KpnI-PstI* fragment isolated from a 4-kb *HindIII* genomic clone (gift from A. Winoto). This probe is immediately upstream of the first J α segment (J α 1) and downstream of ψ J α . It hybridizes to a 1.5-kb *SacI-HindIII* fragment²⁹. Note that m δ Rec- ψ J α rearrangements do not affect the outcome. The C α probe is a 0.7 kb *XbaI-EcoRI* fragment encompassing the second exon of TCR C α and hybridizes to 1.3-kb *SacI-HindIII* fragment. The mutation in TCR C α does not affect the results. The intensity of the radioactivity was measured using a Fujix BAS2000 Bio-image Analyzer within the linear range of intensities. Total RNA was prepared for northern blot analysis. The TCR C α probe was a 550-bp *NcoI* fragment isolated from cDNA 2C1B2 α 4 (gift from Y. Takagaki), which spans the insertion site of pgk-*neo*. The TCR C β 2 probe was a 800-bp genomic *SacI* fragment within the TCR C β 2 region, isolated from cosmid 2.3W7 (ref. 16).

TCR- α rearrangements in thymocytes

Experiment	RAG-1 -/-	TCR- β +/-	TCR- β -/-	Experiment	RAG-1 -/-	TCR- β +/-	TCR- β -/-
(1) Ratio 5'J α 1/C α	7.59	4.35	5.48	(3) Ratio 5'J α 1/C α	3.90; 4.03		2.34; 3.26;
$RI(\alpha)$	0	43	28				3.42
(2) Ratio 5'J α 1/C α	3.44; 3.58	1.90; 2.14;	2.41; 2.67;	Average ratio	3.97		3.01
		2.17; 2.35	2.97; 2.98	$RI(\alpha)$	0		24
Average ratio	3.51	2.14	2.76	(4) Ratio 5'J α 1/C α	8.49; 8.69	5.00	5.35; 5.53;
$RI(\alpha)$	0	39	21				5.97; 6.40;
							6.62
				Average ratio	8.59	5.00	5.97
				$RI(\alpha)$	0	42	30

at least some of the surface TCR- β chains are not associated with any of the other known complete TCR polypeptide chains.

A few CD4-positive lymphocytes became increasingly detectable with age in the peripheral lymphoid organs and in the blood of TCR- α mutant mice but not in TCR- β mutant mice, which stained weakly with antibodies against TCR- β and CD3- ϵ . As an example, staining patterns of lymphocytes from mesen-

teric lymph nodes of a 9-month-old TCR- α mutant mouse and 8-month-old TCR- β mutant mouse are shown in Fig. 6a. The TCR- β and CD4-positive cells were not stained by two different anti-TCR C δ antibodies (3A10 and GL3) and are therefore unlikely to bear mixed TCR $\beta\delta$ heterodimers¹⁷ (data not shown). We observed an equally small population of CD4-positive dull TCR- β -positive cells in the periphery of TCR- β transgenic TCR- α mutant mice (data not shown).

FIG. 5 FACScan of TCR- β transgenic mutant thymocytes; CD3- ϵ -FITC (horizontally) and β -PE (vertically). WT, wild-type; β Tg, TCR- β transgenic; α -/-, TCR- α mutant; RAG-1 -/-, RAG-1 mutant. METHODS. Thymocytes were isolated from 4-week-old mice and stained as described for Fig. 3a.

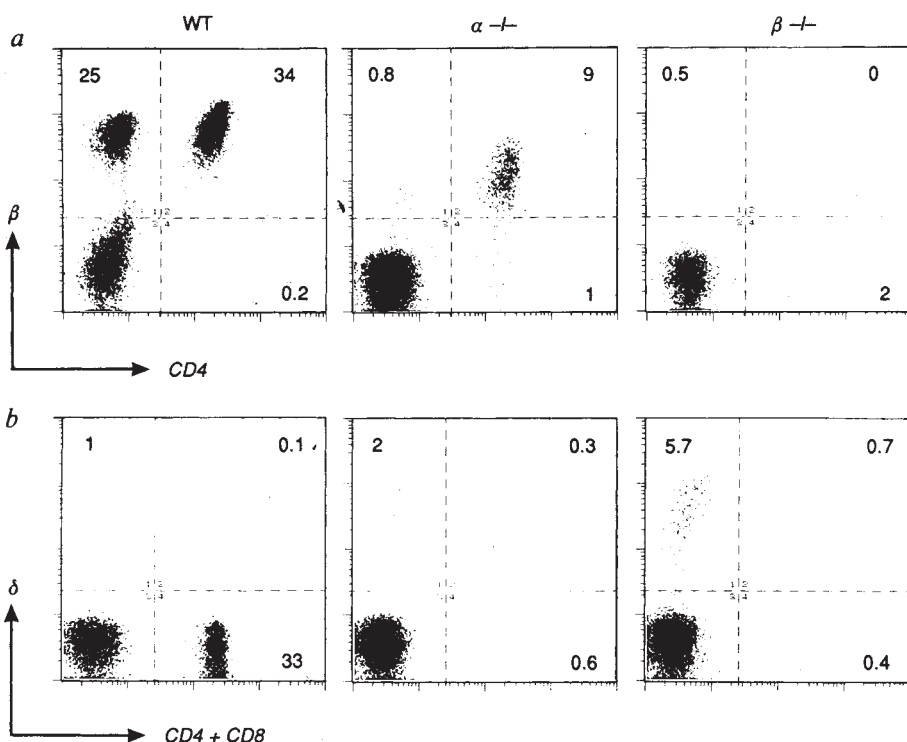
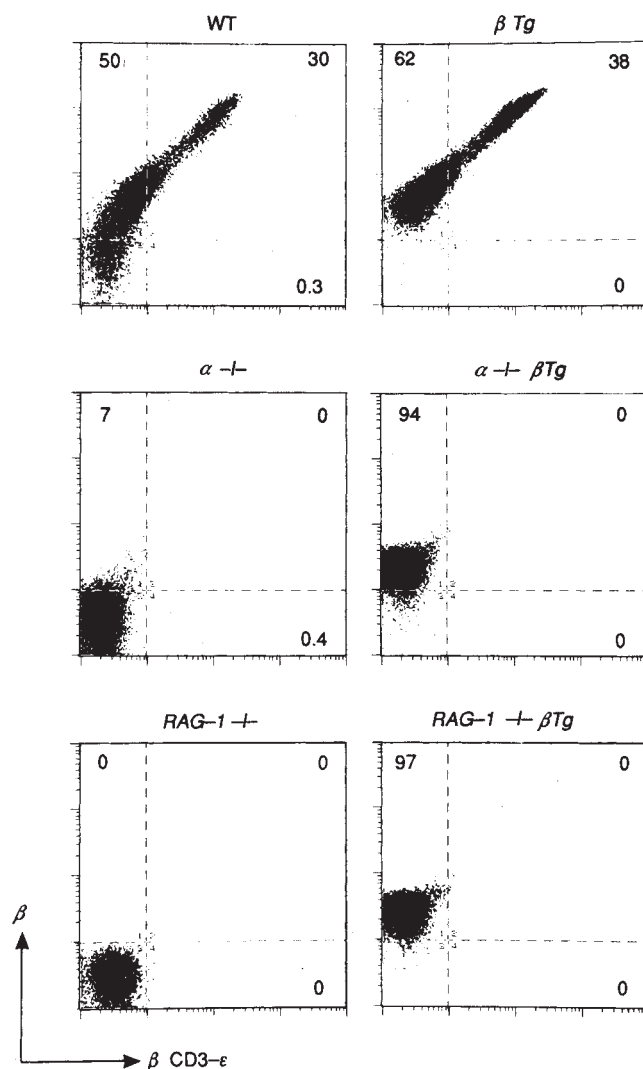


FIG. 6 *a*, Dull TCR- β -positive cells in the periphery of the TCR- α mutant mouse. Lymphocytes are from mesenteric lymph nodes of a 9-month-old wild-type mouse (*left*), a TCR- α mutant littermate (*middle*), and an 8-month-old TCR- β mutant mouse (*right*). CD3- ϵ -FITC (horizontally) and TCR- β -PE (vertically) the total numbers of lymphoid cells were 5×10^6 , 22×10^6 and 9×10^6 , respectively. *b*, $\gamma\delta$ T cells in the periphery of young TCR- α and TCR- β mutant mice. The number of $\gamma\delta$ T cells in the mutant mice is similar or slightly increased compared to the number in wild-type younger littermates (less than 2–3 months old), and most $\gamma\delta$ T cells in the mutant mice are CD4-CD8 double negative. Splenocytes were from a 5-week-old (*left*) TCR- α heterozygous, (*middle*) TCR- α mutant, and (*right*) TCR- β mutant mouse. Staining was with CD8-FITC + CD4-FITC (horizontally) and TCR- δ -PE (vertically). The total number of cells was respectively 44×10^6 , 60×10^6 and 40×10^6 lymphoid cells. METHODS. The mice in each panel were stained in parallel using antibodies described for Fig. 3a. The gates were put on the small, lymphoid-like cells. For both *a* and *b*, the control mouse is a littermate of the TCR- α mutant mouse.

$\gamma\delta$ T-cell development

In both the TCR- α and TCR- β mutant mice, $\gamma\delta$ T-cell development in the thymus appears to be unaltered; the number of $\gamma\delta$ thymocytes is the same as in wild-type mice (Fig. 3b). The numbers of $\gamma\delta$ T cells in the spleen, lymph nodes and gut epithelium are similar or slightly increased in young mutant mice as compared to wild-type littermates (an example for the spleen is shown in Fig. 6b).

Discussion

We have shown that TCR- β rearrangement or, more likely, expression of a functionally rearranged TCR- β gene, is not only necessary but also sufficient for driving most of the immature IL-2R-positive DN thymocytes to the DP stage and for expanding the pool of thymocytes. We have also shown that TCR- α expression is irrelevant to these processes, but together with TCR- β rearrangement and expression it allows further differentiation into mature, CD4 or CD8 SP thymocytes expressing high levels of $\alpha\beta$ TCR and CD3.

The conclusion that TCR- β is sufficient is based on the observations that both the severe drop in total thymocyte numbers (to about 1% of the wild-type level) and the blockade at a DN stage in the RAG-1 mutant mice are fully restored upon introduction of a productively rearranged TCR- β transgene. These observations differ from earlier findings in which *scid* mice¹⁸ were host for the same TCR- β transgene^{8,9}. In these TCR- β transgenic *scid* mice the number of double-positive cells reaches only about 2% of wild-type level and the total number of thymocytes is about 3% of wild-type⁸. The most likely explanation for the difference is that the *scid* mutation affects the physiology of the thymocyte in a more general way rather than just preventing correct *V(D)J* recombination¹⁸.

The origin of the small number (about 5% of the wild-type level) of DP cells remaining in the thymus of TCR- β mutant mice is unclear. They may belong to a minor $\alpha\beta$ T-cell pathway in which DN $\rightarrow$ DP transition occurs independently of TCR- β rearrangement or expression. Alternatively, these cells may belong to the $\gamma\delta$ T-cell lineage because they are virtually abolished when TCR- δ mutation is introduced into the TCR- β mutant mice. No role for DP cells as intermediates in the differentiation of $\gamma\delta$ T cells has been demonstrated. But TCR- δ -

positive DP cells have been detected in the late embryonic thymus³ and about 1% of TCR- δ -positive cells in the postnatal TCR- β mutant thymus are DP (data not shown).

The hypothesis that TCR- β rearrangement provides a signal for TCR- α rearrangement is challenged by our findings of TCR- α rearrangements and full-size TCR- α transcripts in the TCR- β mutant mice. These findings are analogous to those made recently in the B-cell system: Ig- κ rearrangements can occur in the absence of an Ig- μ product in Ig-C μ mutant mice¹⁹. However, rearrangements may occur sequentially at the TCR- β and TCR- α loci, but with a causal relationship.

Surface expression of TCR- β without TCR- α (α -less TCR) has been found in DN thymocytes of TCR- β transgenic mice²⁰, in DN and DP thymocytes of TCR- β transgenic *scid* mice^{8,9}, and in a DP thymocyte line²¹. We have now extended these observations to DP thymocytes in TCR- β transgenic TCR- α mutant, TCR- β transgenic RAG-1 mutant and probably to TCR- α mutant mice. The α -less TCR could be an artefact of the mutant or TCR- β transgenic mutant mice. But cells expressing such TCR may be normal intermediates of $\alpha\beta$ thymocyte differentiation, in analogy to the pre-B cells that expresses Ig- μ without any light chains on the surface. Surface expression rather than only cytoplasmic expression of the TCR- β chain may be essential for transduction of the signal for the induction of CD4 and CD8 and/or proliferation of DP cells. This hypothesis can be tested by crossing variously altered TCR- β transgenes¹⁴ into the RAG-1 or TCR- β mutant mice.

The presence of CD4-positive dull TCR- β positive cells in the periphery of the TCR- α mutant and TCR- β transgenic TCR- α mutant mice, suggests that some thymocytes bearing TCR- β without TCR- α may develop into mature T cells and emigrate to the periphery in normal mice, although such T cells could be present only in the genetically manipulated mice. Recently another strain of TCR- α mutant mice has been reported²² but no T cells with α -less TCR were observed; the reason for this discrepancy is not clear, although these cells may have been overlooked.

Finally, our studies indicate that the TCR- α or TCR- β mutation does not block $\gamma\delta$ T-cell development. The $\alpha\beta$ T-cell-deficient mice should be useful for analysing the *in vivo* functions of $\gamma\delta$ T cells. □

Received 2 July; accepted 20 October 1992.

1. Tonegawa, S. *Nature* **302**, 575–581 (1983).
2. Davis, M. M. & Bjorkman, P. J. *Nature* **334**, 395–402 (1988).
3. Itohara, S., Nakanishi, N., Kanagawa, O., Kubo, R. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5094–5098 (1989).
4. Haas, W. & Tonegawa, S. *Curr. Opin. Immun.* **4**, 147–155 (1992).
5. Shortman, K. *Curr. Opin. Immun.* **4**, 140–146 (1992).
6. Raulet, D. H., Garman, R. D., Saito, H. & Tonegawa, S. *Nature* **314**, 103–107 (1985).
7. Snodgrass, R. H., Dembic, Z., Steinmetz, M. & von Boehmer, H. *Nature* **315**, 232–233 (1985).
8. von Boehmer, H. *A. Rev. Immun.* **8**, 531–556 (1990).
9. Kishi, H. *et al. EMBO J.* **10**, 93–100 (1991).
10. Mombaerts, P. *et al. Cell* **68**, 869–877 (1992).
11. Mombaerts, P., Clarke, A. R., Hooper, M. L. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3084–3087 (1991).
12. Gossler, A., Doetschman, T., Korn, R., Serfling, E. & Kemler, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9065–9069 (1986).
13. Uematsu, U. *et al. Cell* **52**, 831–841 (1988).
14. Krimpenfort, P., Ossendorp, F., Borst, J., Melief, C. & Berns, A. *Nature* **341**, 742–746 (1989).
15. Kubo, R. T., Born, W., Kappler, J. W., Marrack, P. & Pigeon, M. *J. Immun.* **142**, 2736–2741 (1989).
16. Malissen, M. *et al. Cell* **37**, 1101–1110 (1984).

17. Hochstenbach, F. & Brenner, M. B. *Nature* **340**, 562–565 (1989).
18. Bosma, M. J. & Carroll, A. M. *A. Rev. Immun.* **9**, 323–350 (1991).
19. Kitamura, D. & Rajewsky, K. *Nature* **356**, 154–156 (1992).
20. von Boehmer, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 9729–9732 (1988).
21. Punt, J. A. *et al. J. exp. med.* **174**, 775–783 (1989).
22. Philpott, K. *et al. Science* **256**, 1448–1452 (1992).
23. Hayday, A. C. *et al. Nature* **316**, 828–832 (1985).
24. Adra, C. N., Boer, P. H. & McBurney, M. W. *Gene* **60**, 65–74 (1987).
25. McBurney, M. W. *et al. Nucleic Acids Res.* **19**, 5755–5761 (1991).
26. Winoto, A. & Baltimore, D. *Cell* **59**, 649–655 (1989).
27. Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 113–151 (IRL, Oxford, 1987).
28. Laird, P. W. *et al. Nucleic Acids Res.* **19**, 293 (1991).
29. Hockett, R. D., Nunez, G. & Korsmeyer, S. J. *New Biol.* **1**, 266–274 (1989).

ACKNOWLEDGEMENTS. We thank R. Kemler, P. Pereira, K. Sano, A. Winoto and A. Berns for reagents. P.M. was a Howard Hughes Medical Institute Predoctoral Fellow, M.A.R. was a Postdoctoral Fellow of the National Cancer Institute of Canada. J.I. is supported by an NIH postdoctoral training grant, and Y.I. is supported by Kao Corporation. This work was funded by grants from the NIH, the Howard Hughes Medical Institute and the Human Frontiers Science Program.

Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2

E. Yvonne Jones*, Simon J. Davis†, Alan F. Williams†, Karl Harlos* & David I. Stuart*

* Laboratory of Molecular Biophysics, The Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK

† MRC Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK

The crystal structure of a soluble form of the T lymphocyte antigen CD2 provides the first complete view of the extracellular region of a cell adhesion molecule. The topology of the molecule, which comprises two immunoglobulin-like domains, is the same as that of the first two domains of CD4 but the relative domain orientation is altered by a fairly flexible linker region. The putative ligand-binding β -sheet forms a flat surface towards the top of the molecule. Crystal contacts between these surfaces suggest a plausible model for the adhesive interaction.

THE four principal families of mammalian cell adhesion molecules are the integrins¹, the cadherins², the selectins³ and the immunoglobulin superfamily (IgSF)⁴. These molecules are involved in low-affinity homophilic and/or heterophilic recognition, but little is known of the physical basis for these interactions.

The IgSF molecule CD2, whose expression is largely restricted to the T-lymphocyte lineage, plays a key part in mediating adhesion between human T lymphocytes and accessory or target cells (reviewed in ref. 5). Originally, the lymphocyte function associated LFA-3 (CD58) glycoprotein was identified as the CD2 ligand^{6,7}; these were the first two molecules shown to be involved in heterophilic adhesion and their interaction is likely to be a paradigm for adhesion interactions within the IgSF. An additional ligand, CD59, the complement-inhibitory glycoprotein^{8,9}, has recently been identified. Attempts to identify a murine homologue of LFA-3 led to the unexpected finding that mouse CD2 interacts with a different molecule, BCM1 (CD48)⁴⁹. In rats the equivalent adhesion interaction is between MRC OX-45 antigen (rat BCM1)¹⁰ and CD2 (P.A. van der Merwe *et al.*, manuscript submitted). An additional role for CD2, involving the cytoplasmic domain in the delivery of mitogenic signals, has been proposed¹¹⁻¹³, but recent experiments¹⁴ indicate a dependence on the T-cell antigen receptor (TCR) complex of T cells, with which CD2, at least transiently, physically interacts¹⁵.

CD2, LFA-3 and BCM1 form a subgroup of the IgSF; the extracellular regions of these molecules are predicted to consist of two IgSF domains, the first of which lacks the canonical disulphide bond¹⁰. A low-resolution NMR structure has confirmed this prediction for rat CD2 domain 1 (ref. 16). All of the LFA-3 binding reactivity resides in domain 1 of human CD2 (ref. 17) and mapping by mutagenesis¹⁸, when interpreted in the context of the NMR structure, indicates that the binding face consists of the β -sheet formed by strands GFCC'C" (ref. 16). The affinity of the monomeric interaction between CD2 and LFA-3 is $\sim 2 \times 10^6 \text{ M}^{-1}$ (ref. 19), somewhat higher than that for other cell-surface interactions^{20,21}. The chromosomal organization of the genes encoding human CD2, LFA-3 and BCM1 is consistent with each gene having evolved by duplication from a common precursor^{22,23}, possibly one encoding a two-domain molecule involved in homophilic adhesion⁴. It has been postulated²⁴ that the 2-fold rotational symmetry of this homotypic interaction is preserved in evolution for the heterotypic interaction between CD2 and LFA-3. Thus, by analogy with the dimerization of antibody variable domains, the CD2/LFA-3 interaction is predicted to involve equivalent β -sheets packing with pseudo 2-fold rotational symmetry²⁴.

We report here the structure of the soluble, extracellular, region of rat CD2 (sCD2) determined at 2.8 Å resolution for two crystallographically independent copies of the molecule. Crystals are held together by homophilic adhesion and the most striking such contact in the sCD2 crystals involves two molecules packed, putative ligand binding-face to binding-face, with 2-fold rotational symmetry. Thus we also report a possible structural paradigm for IgSF adhesive interactions.

Structure determination

The production and crystallization of sCD2 (residues 1 to 177) is described in ref. 25. sCD2 has four potential N-linked glycosylation sites. As glycosylation is generally detrimental to crystallization²⁶, sCD2 was expressed in a lectin-resistant mutant of the Chinese hamster ovary cell line, Lec3.2.8.1 (ref. 27), resulting in a secreted glycoprotein with oligosaccharides (predominantly Man₅GlcNAc₂) that are amenable to deglycosylation by endoglycosidase H (ref. 25). After purification by immunoaffinity chromatography and gel filtration, the deglycosylated sCD2 readily crystallized. The crystals belong to space group *P*4₁2₁2, with unit cell dimensions *a* = *b* = 111.4, *c* = 86.9 Å and diffract to Bragg spacings of 2.5 Å when exposed to synchrotron radiation.

Details of the structure determination to 2.8 Å resolution are given in Table 1. The crystals contain two sCD2 molecules per asymmetric unit (60% crystal solvent content). Multiple isomorphous replacement (MIR), phase improvement and extension by solvent flattening²⁸ and, finally, a single round of 2-fold density averaging²⁹ (D.I.S. and J. Grimes, unpublished) yielded unambiguous electron density (Fig. 1a). The resultant models for the two copies of sCD2 were refined in the program XPLOR³⁰. Clearly there was a relative reorientation between the two domains in copies one and two (i.e. hinge-bending). The domains were therefore separately restrained to obey the non-crystallographic symmetry during refinement.

The *R* factor for the current structure (residues 2 to 176) is 21.5% for all data to 2.8 Å (19.6% for *F* > 3 σ) with a bulk solvent correction but no model for the ordered solvent or sugars (Fig. 1b). A test refinement (coordinates not subsequently used), with non-crystallographic symmetry restraints switched off, indicates the accuracy of the atomic positions to be 0.3 Å (by analysis of the 81% of the molecule considered to be unperturbed by crystal packing forces; Table 1). The quality of the electron density is sufficient to determine reliable residue positions throughout (*B*-factors are presented in Table 1) but before residue 5 the electron density clearly indicates two possible conformations for the N terminus, and at present only one has been modelled.

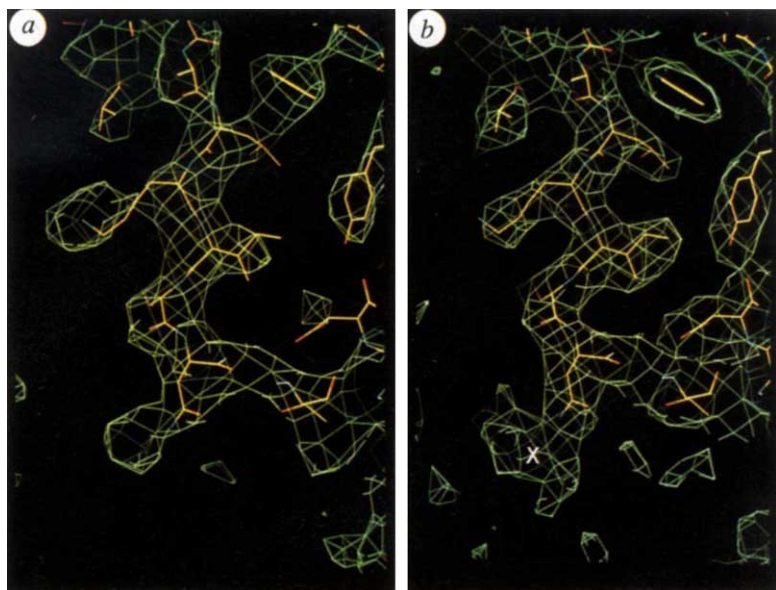


FIG. 1. Portions of electron density maps for residues 63–67 of sCD2 copy 1. *a*, 4 Å solvent-flattened MIR map. *b*, $2F_{\text{obs}} - F_{\text{calc}}$ map (model phases) after refinement at 2.8 Å resolution. Density (indicated by an X in *b*) is clearly visible at N67 for an *N*-linked acetyl glucosamine which was not included in the phasing model. Similar quality density is present at N112 in both copies and, by virtue of stabilizing lattice contacts, at N84 in copy 1. Of the expected *N*-linked sites, only N84 in copy 2, in a region of high *B* factors, lacks any trace of sugar density.

Molecular structure

The extracellular region of CD2, as predicted¹⁰, consists of two antiparallel β -barrels with immunoglobulin folds. So far the most comparable structure is that of the NH₂-terminal two-domain fragment of CD4 (D1D2) (refs 31, 32). A structure-based sequence alignment is given in Fig. 2. In sCD2, as in CD4(D1D2), the domains abutt to give a rod-like molecule, which for sCD2 has approximate dimensions of $20 \times 25 \times 75 \text{ \AA}$. But in marked contrast to CD4(D1D2), there is a linker region (residues 99 to 103) between the two domains. When considered in the context of the cell surface, the overall structural arrangement is likely to result in the domain 1 GFCC'C'' sheet forming a highly accessible platform some $75 \text{ \AA}$ above the cell surface (Fig. 3). The structure naturally divides into four regions: two IgSF domains, the interdomain linker region, and an extended C-terminal tail.

Domain 1. The crystal structure for sCD2 domain 1 is in close agreement with the polypeptide fold of the single domain in solution as determined by NMR spectroscopy¹⁶ (a detailed comparison will be reported elsewhere). The topology is that of an immunoglobulin variable (V) domain with two β -sheets of

strands conventionally labelled DEB and AGFCC'C" (Fig. 3a). Superpositions with domain 1 of CD4 and the V κ -domain of the Bence-Jones protein REI (ref. 33) are shown in Fig. 4a and b. Reference may also be usefully made to the recently determined structure of a soluble form of CD8 (ref. 34). Despite the absence in CD2 of the usual disulphide, the positions of all the β -strands in the V domain framework are highly conserved. However, of the loop regions only AB and EF have similar conformations in CD2, CD4 and REI.

Loop regions CC', C'C" and FG bound the functionally important GFCC'C" sheet. As in CD4, the C'C" loop is longer than in the classic V domain, represented here by RE1, but unlike CD4, the orientation maintains the flat surface of the sheet. A similar C'C" loop conformation is observed in CD8 (ref. 34). The CC' and FG loops mediate dimerization in antibody V domains³⁵ and CD8 (ref. 34). Both are shortened in CD4 (which does not dimerize)^{31,32} relative to the classic V domain^{31,32} but in sCD2 the CC' loop is shortened and the FG loop is greatly lengthened. β -Bulges in the C' and G strands, which introduce a distinctive twist to these strands, are believed to be crucial to V domain dimerization³⁵. Although the conserved

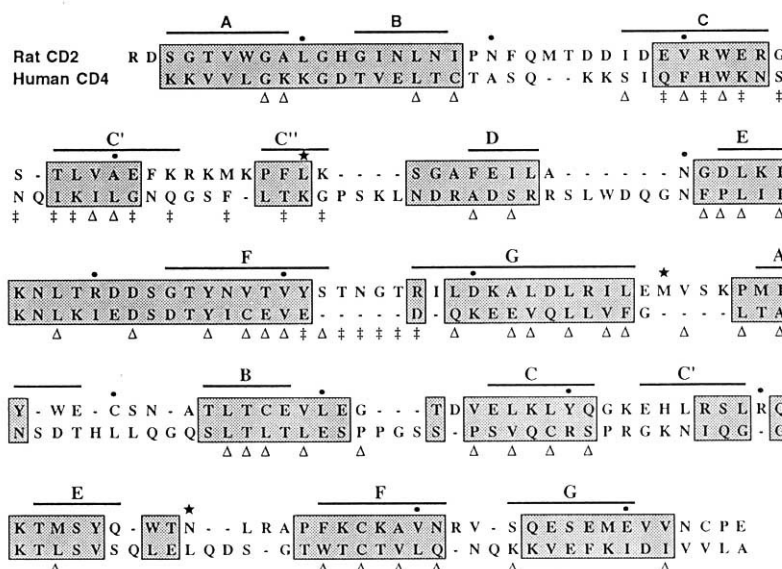


FIG. 2. Sequence alignment of rat CD2 and human CD4. The amino-acid sequences were aligned on the basis of structural superpositions. Black dots indicate each tenth residue and stars every fiftieth residue of the rat CD2 sequence. The principal differences from previous alignments occur at β -strand A of domain 2. Residues judged to be structurally equivalent in terms of α -carbon position and side-chain direction in the superimposed structures are stippled. Positions of the β -strands in sCD2 (based on program DSSP⁴⁶) are indicated and solvent inaccessible residues (less than 20 Å² of exposed surface) are indicated by triangles. Residues mediating the 2-fold lattice contact are indicated by the symbol ‡.

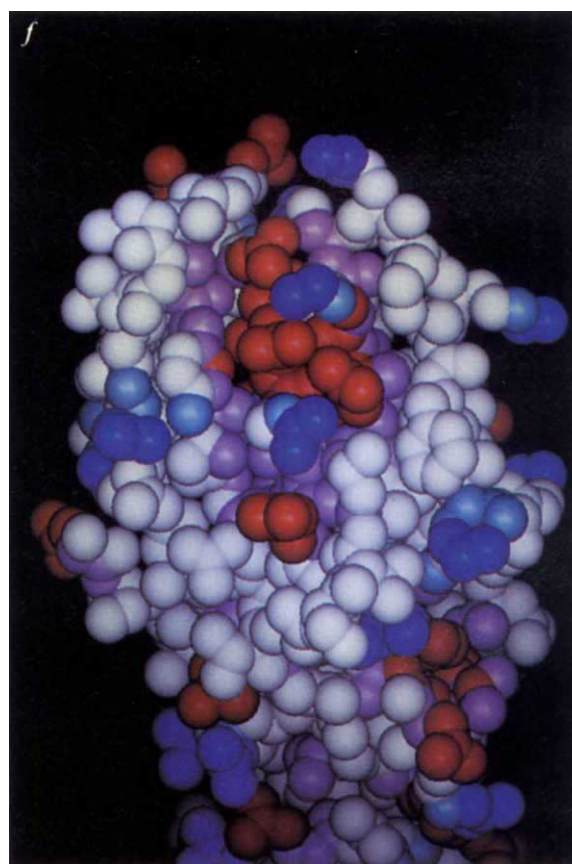
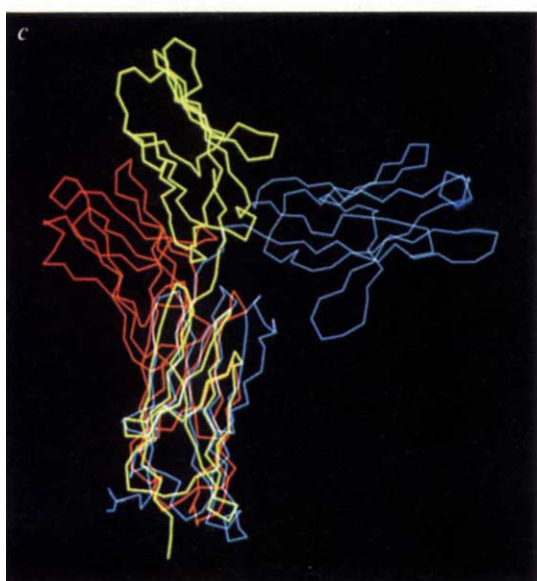
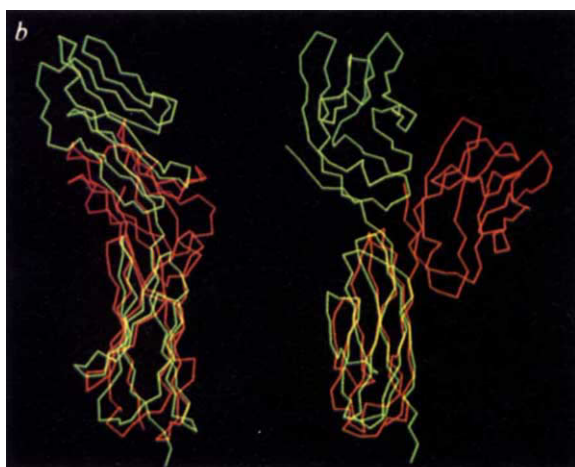
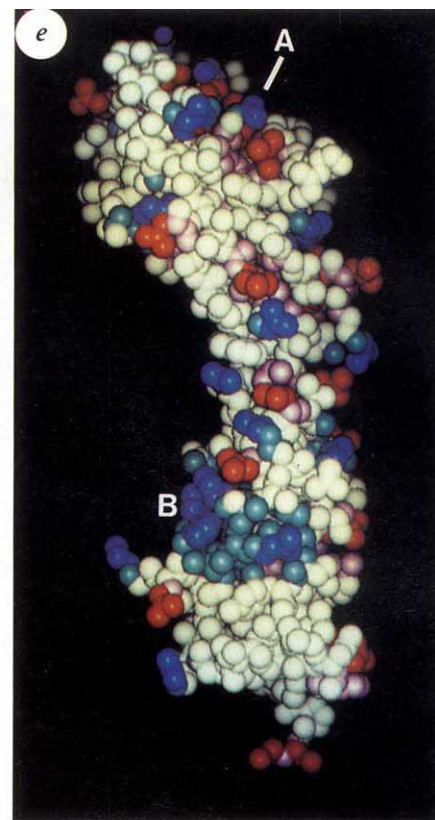
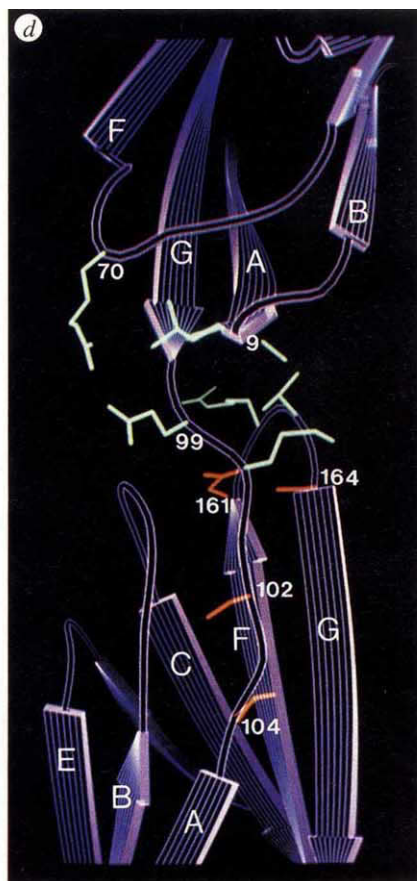
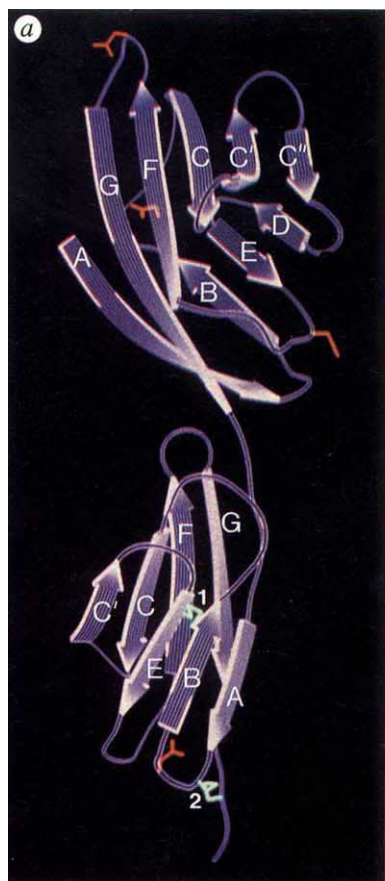


FIG. 3 The structure of sCD2. The orientation of sCD2 is chosen relative to a cell surface assumed to lie at the bottom of the illustration in the plane normal to the vertical. *a*, Schematic diagram of the polypeptide fold (program RIBBON⁴⁷). Several of the loops in domain 1 are flexible; r.m.s. deviations between main-chain atoms for the two copies of sCD2 exceed 1.0 Å for residues 20–26, 34–37, 44–47, 70–71 and 85–86. Bifurcation of the electron density at Gly4 indicates a degree of structural fluidity in the first 4 residues of the polypeptide chain. The alternative conformation extends away from the molecule, stabilized by crystal contacts. The disulphide bonds are indicated in green (1) Cys 117 and 157; (2) Cys 110 and 174. Asparagine residues providing *N*-linked oligosaccharide sites are indicated in red. N67 lies on the EF loop at the bottom of domain 1, N77 in strand F is orientated towards the bottom of the F and G strands. N84 at the tip of the FG loop points back, above the BC loop. N112 lies on the AB loop at the base of domain 2. *b*, α -Carbon trace superpositions for CD2 and CD4. Two orthogonal views are shown. The molecules are matched on CD2 domain 2. CD2, green; CD4, red. *c*, α -Carbon trace superpositions for CD2, CD4 and Fab NEW light chain. The molecules are matched on CD2 domain 2. Note the Fab domain 1 is flipped over by some 90° with respect to CD2 domain 1. CD2, green; CD4, red and NEW, blue. *d*, Linker region. Residues involved in hydrophobic and van der Waals interactions at the domain interface are indicated in green and include W7, A9, L10, E99, M100, R162 and V163, with R70 forming a salt bridge to E99. This type of link may be characterized by the stretch of apolar residues in the N-terminal β -strand A and a hydrogen bond network in domain 2 mediated by Ser-X-Pro at the start of β -strand A combined with Asn 161 and Ser 164 at the end of β -strand F and start of G (residues in red). *e*, Solid sphere representation (van der Waals surface). The colour coding accords with the electrostatic potential computed with DELPHI⁴⁸ (in collaboration with C. Edge) at neutral pH; blue represents positive potential, red negative, and white neutral. The orientation is almost orthogonal to that in *a*. The surface of the domain 1 GFCC'C" sheet is indicated by 'A' and the domain 2 GFCC'C" sheet by 'B'. *f*, Domain 1 GFCC'C" sheet. Colour coding as in *e*. The view is along the normal to the β -sheet surface as indicated by the line in *e*.

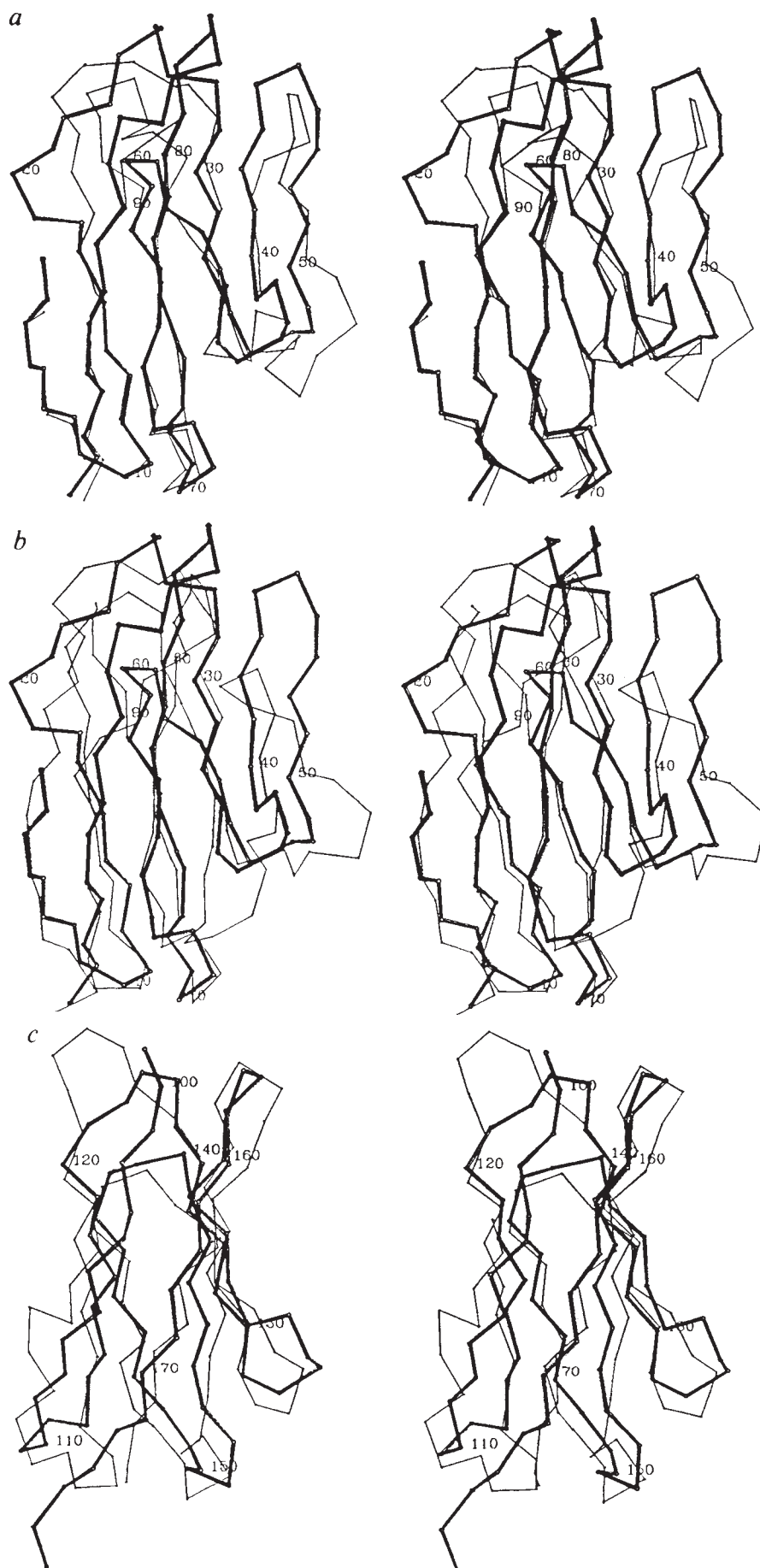


FIG. 4 Stereodiagrams of pairwise α -carbon backbone superpositions. *a*, CD2 domain 1 (thick lines; residues 2–98) and CD4 domain 1 (residues 1–98) based on 70 equivalenced $C\alpha$ atoms fitted with 0.83 Å r.m.s. deviation. *b*, CD2 domain 1 (thick lines) and REI (residues 1–114) based on 60 equivalenced $C\alpha$ atoms fitted with 0.82 Å r.m.s. deviation. *c*, CD2 domain 2 (thick lines; residues 99–176) and CD4 domain 2 (residues 99–176) based on 58 equivalenced $C\alpha$ atoms fitted with 1.69 Å r.m.s. deviation. (Superpositions were done using the program SHP⁴².)

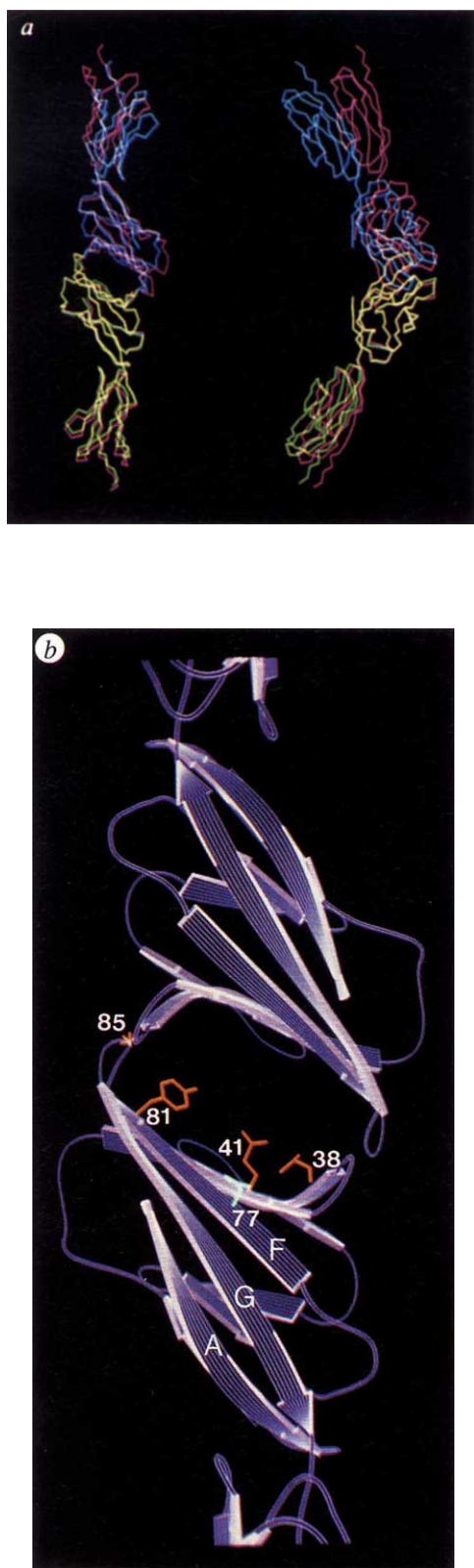


FIG. 5 Homophilic interaction mode observed in crystal. *a*, α -Carbon trace superposition of the two versions of this interaction. Copy 1 molecules are coloured green and blue, copy 2 molecules red. Superposition is on domain 1 of the blue molecule. The hinge-bending variation between the two copies may also be noted. *b*, Close-up of the molecular association at the crystallographic 2-fold axis. Residues implicated in human CD2/LFA-3 and CD2/CD59 binding are indicated and those also involved in the homophilic interaction are colour-coded red.

immunoglobulin β -bulge sequence motifs are absent, a corresponding pair of β -bulges occurs in CD2 but these features are now positioned immediately after the CC' and FG β -turns (at Leu 38 and Arg 87 respectively). Thus the characteristic twist is limited to the tips of these loop regions (Fig. 4b).

Linker region. The participation of β -strand G in the domain 1 GFCC'C'' β -sheet ends at Leu 98. The main chain continues in an extended conformation held at the edge of the domain 2 β -sandwich by hydrogen bonds to strands from both β -sheets until kinked at proline 104 to commence the standard β -strand A of domain 2. The overall effect is to loosen the tightly abutting domain junction seen in CD4(D1D2) by sliding domain 2 down by 4 residues with respect to domain 1, lengthening the molecule by some 15 Å. The run of four buried apolar residues which distinguishes the G-to-A strand transition in CD4 is absent. The interface is stabilized by some 400 Å² of total buried surface area, much less than the 880 Å² buried in the more intimate association of CD4(D1D2).

As in CD4(D1D2), domains 1 and 2 of sCD2 are related by a rotation of roughly 160° about the axis of the rod-like molecule (the vertical in Fig. 3b). But as shown in Fig. 3b for the sCD2 and CD4(D1D2) structures superimposed with respect to their second domains, a tilt of some 45° from this rod axis is required to superimpose their first domains. This reorientation eliminates the 'notch', a structural feature in CD4 that may mediate interaction with the class II major histocompatibility complex (MHC) molecule^{31,36}. Both CD2 and CD4 have very different relative domain orientations to those of immunoglobulins (Fig. 3c). Slight changes in main-chain torsion angles for the linker residues in sCD2 can generate substantial 'hinge-bending' shifts in relative domain orientation. A 7° hinge-bending effect is observed on comparison of the two independent copies and a similar or greater range of flexion may be expected *in vivo*. This linker region establishes a second class for connections between unpaired immunoglobulin domains, the characteristics of which are shown in Fig. 3d.

Domain 2. The topology of domain 2 has both V-like and C-like features. It most closely resembles domain 2 in CD4 (Fig. 4c), with a β -barrel comprising sheets EBA and GFCC'. However, although topologically equivalent, the core strands for CD2 and CD4 cannot be precisely superimposed (the r.m.s. deviation of 1.7 Å for 58 equivalenced C α atoms being twice as much as that between domain 1 cores; see Fig. 4). Indeed, a classic C-like domain (C_L from Fab NEW³⁷) superimposes on CD2 equally well (r.m.s. deviation of 1.7 Å, based on 59 equivalenced C α atoms) because the intersheet spacing is more closely matched. The two disulphide bonds are as predicted¹⁰. The bond between Cys 117 and Cys 157 corresponds to the standard immunoglobulin disulphide, whereas cysteines 110 and 174, linking the ends of the A and G strands, tether the bulk of the molecule to the C-terminal stalk. CD4 has neither of these bonds, which allows an expansion of the intersheet spacing compared to that of CD2 (Fig. 4c). Most of the loops adopt similar conformations in CD2 and CD4, the exceptions being the AB and BC loops.

A superposition of CD2 domains 1 and 2 reveals a striking match on the GFCC' strands (r.m.s. deviation of 0.79 Å for 35 equivalenced C α atoms), emphasizing the partial V-like character of domain 2. The CC' loops correspond precisely, but the FG loop in domain 2 lacks the β -bulge and accompanying twist seen in domain 1. The surface of this β -sheet is predominantly composed of charged side chains in both domains, but in domain 2 these are almost exclusively lysine and arginine residues which results in a very pronounced patch of basic potential (Fig. 3e).

C-terminal stalk region. Glutamic acid 170 forms the final main-chain hydrogen bond in the molecule. The remaining 7 C-terminal residues of sCD2 adopt an extended conformation. Valine 172 and the disulphide bridge residues 110 and 174 contribute to the hydrophobic core of domain 2, after which the path of the main chain is locked to continue on into solution by Pro 175. The base of domain 2 is neutrally charged (Fig. 3e),

TABLE 1 Statistics for crystallographic structure determination

Diffraction data					
Data set	$d_{\min}$ (Å)	No. of measurements	Unique reflections (% complete)	R_{merge} (all data)	MFID
Native (IP Lab)	4.0	10,144	3,936 (79)	11.4	—
Native (IP SRS)	2.8	51,452	13,432 (96)	12.6	—
EMP	4.0	10,047	4,371 (88)	13.9	17.5
NaAuCl ₄	4.0	10,279	4,314 (87)	9.4	25.7
K ₂ HgI ₄	4.0	8,020	3,928 (79)	11.7	18.9
Baker's	4.0	10,213	4,345 (87)	8.7	26.8
Pb(OOCCH ₃) ₂	4.0	10,327	4,152 (83)	8.1	22.1

MIR phasing			
Derivative	No. of sites	R_c to 4.5 Å (4.0 Å)	Phasing power to 4.5 Å (4.0 Å)
EMP	3	0.65	0.8
NaAuCl ₄	4	0.61 (0.63)	1.5 (1.5)
K ₂ HgI ₄	1	0.70	0.9
Baker's	3	0.64	1.5
Pb(OOCCH ₃) ₂	2	0.59 (0.66)	1.7 (1.1)

Solvent flattening				
$d_{\min}$ (Å)	Pre-Wang FM	(%) Solvent	Wang R factor	Phase change
4.5	0.57	60	23.4	26.4°
4.0	0.54	60 (50)	26.1* (18.9)	46.7° (52.0°)
3.0	—	60 (55)	19.2 (16.8)	—

Refinement: non-crystallographic symmetry (n.c.s.) restraints			
Domain	R.m.s. deviation in position	R.m.s. deviation in B	Residues restrained by n.c.s.
1	0.29 Å	0.46 Å ²	4:18 28:33 39:41 49:68
2	0.27 Å	0.42 Å ²	73:83 88:98
			99:174

Crystallization and preparation of derivatives. Crystals were grown at 4 °C in microbridges⁴¹ by equilibration against reservoirs of 18% PEG 4000, 50 mM Tris, pH 8.0, starting from sitting drops composed of 4 μ l of protein at 17 mg ml⁻¹ in 25 mM Tris, 140 mM NaCl, pH 7.4, and 2 μ l reservoir²⁵. Tests of some 40 heavy-atom soak conditions at 4 °C yielded five useful derivatives: (1) ethyl mercury phosphate (EMP): 5 mM in 18% PEG 4000, 50 mM Tris pH 8.0 for 16 h; (2) NaAuCl₄: 1 mM in 18% PEG 4000, 25 mM Tris, 140 mM NaCl, pH 7.4, for 5 h; (3) K₂HgI₄: 10 mM in 50 mM KI, 18% PEG 4000, 25 mM Tris, 140 mM NaCl, pH 7.4, for 18 h; (4) 1,4-diacetoxymethyl-2,3-dimethoxybutane (Baker's dimercurial); 20 mM (saturated) in 18% PEG 4000, 50 mM Tris, pH 8.0, for 68 h; (5) Pb(OOCCH₃)₂: 10 mM (saturated) in 18% PEG 4000, 50 mM Tris, pH 8.0, for 66 h. **Diffraction data.** Crystals with typical dimensions of 0.08 × 0.08 × 0.3 mm gave relatively weak diffraction. IP(SRS) native data were collected using a Mar-research imaging plate (IP) system (J. Hendrix and A. Lentfer) on station 9.5 (wavelength 0.90 Å) at the SERC Synchrotron Radiation Source (SRS), Daresbury. IP(lab) native and all derivative data were measured with CuK α X-rays generated by a Rigaku RU200 rotating anode with graphite monochromator and detected on an IP system. IP data were processed using the MOSFLM program package (A. Leslie); native and derivative data were scaled using programs 3DSCALE and DIFFER⁴². Additional data collected inhouse on a Siemens/Xenotronics area detector and at the SRS on an Enraf-Nonius FAST area detector (detailed in ref. 25) were merged with the IP SRS data for some of the stages of the structure determination but refinement and the current model are based solely on the IP SRS data. ($R_{\text{merge}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$; MFID, mean fractional isomorphous difference). **MIR analysis.** The EMP derivative was solved by the automatic Patterson search program GROPAT^{43,44}; the others by difference Fourier syntheses and inspection of difference Patterson maps. The EMP, NaAuCl₄ and Baker's dimercurial derivatives have two major sites in common which conform to the non-crystallographic symmetry (binding at His 12). Heavy-atom parameters were refined and MIR phases calculated using the programs REFIN and PHASE⁴⁴ (P. Shaw, unpublished). (R_c , Cullis R factor for centric reflections; phasing power, mean value of the heavy-atom structure factor amplitudes divided by the residual lack-of-closure error.) **Solvent flattening and phase extension.** MIR phases based on all 5 derivatives were refined at 4.5 Å resolution in the Wang program suite²⁸ by 7 cycles of solvent flattening following each of 3 envelope determinations (60% solvent). In an iterative procedure⁴⁴, commencing at 4.6 Å resolution, MIR (NaAuCl₄ and Pb(OOCCH₃)₂) and calculated phases were included to 4.0 Å and 3.0 Å resolution respectively. This phase extension was achieved in continuous steps of 0.0022 Å⁻¹, with envelope redeterminations in every 7th cycle. Fourteen cycles of solvent flattening with 60% solvent, followed by 21 with 55%, yielded the final 3 Å map, and a further 21 cycles with 50% solvent, yielded a definitive 4 Å map. The 3 Å map was noisier but showed improved definition of side-chain electron density. The non-crystallographic symmetry was established from the electron density and heavy atom sites. A single round of real space density-averaging using the GAP program suite (DIS and J. Grimes) improved main-chain connectivity in both 4 Å and 3 Å maps. (FM, mean figure of merit; Wang R factor, the crystallographic R factor between observed structure factor amplitudes and those calculated from the solvent flattened map; phase change, differences between MIR phases and phases from the solvent flattened map; asterisk, not converged.) **Model building and refinement.** The polypeptide chain was fitted to unambiguous electron density (Fig. 1a) for one copy of the molecule using the program FRODO⁴⁵ with reference to four maps: solvent-flattened and density-averaged at 4 Å and 3 Å resolution. The NMR solution structure for domain 1 (ref. 16) could be placed immediately in the electron density to give a very good fit. This was optimized by simple rigid-body manipulations of some of the loop regions. The crystal structure for CD4 domain 2 (refs 31, 32), automatically modified (program MUTATE, R. Esnouf) to accord with the predicted CD2/CD4 sequence alignment¹⁶, provided a putative model for sCD2 domain 2; major rebuilding was required to match this to the electron density. The atomic coordinates for the second copy of sCD2 were generated by non-crystallographic symmetry. The two models were initially refined at 3.0 Å, using simulated annealing, in the program XPLOR³⁰, and the extent of the non-crystallographic symmetry established. Five rounds of model rebuilding and refinement, during which the resolution was extended to 2.8 Å, resulted in the current models which comprise a total of 2,796 non-hydrogen atoms. All 13,432 reflections measured to spacings of 2.8 Å are included in the refinement. The stereochemistry is typified by an r.m.s. bond deviation from ideality of 0.015 Å and B factors restrained such that the r.m.s. discrepancy in values between bonded atoms is 4.8 Å². Mean B factors of 29.5 Å² and 37.1 Å² for domain 1 in copies 1 and 2 contrast with values of 23.5 Å² and 23.3 Å² for domain 2. The thermal parameters for main-chain atoms exceed 50 Å² for residues 2–4 and 175–176 in both copies and for residues 19–26, 43–48 and 84 in copy 2. For these models, the R factor for all data to 2.8 Å is 21.5% (24.3% if tighter B factor restraints, typified by 1.0 Å² r.m.s. discrepancy between all bonded atoms, are applied). All non-glycine ϕ , ψ angles lie in allowed regions of the Ramachandran plot.

thus there is no obvious suggestion of an interaction with the cell surface. The predicted amino-acid sequence for CD2 (ref. 38) suggests that a hydrophobic transmembrane helix starts after Pro 180. If the polypeptide chain is assumed to continue in an extended conformation from Pro 175 to Pro 180, it will span some 16 Å. As the polar headgroups of the lipid bilayer will account for some 6 to 10 Å of this distance, the overall stalk length, from the base of domain 2 to the cell surface, is likely to be less than 10 Å. The implications of this limitation for the relationship between CD2 and the cell surface will be considered elsewhere.

Glycosylation

sCD2 contains four potential *N*-linked glycosylation sites³⁸ (asparagines at positions 67, 77, 84 and 112), which after endoglycosidase H digestion retain single *N*-acetyl glucosamine moieties. Clear electron density for the residual sugar is present at each of these residues in at least one of the two independent copies of the molecule (Fig. 1), indicating the orientations of the roots of the oligosaccharides on this glycoprotein (Fig. 3a). The three domain-1 oligosaccharides lie on the perimeter of the β -sheet believed to be involved in the adhesive interaction, leaving the face of the sheet relatively free for protein-protein interactions. The only site conserved between rat and human CD2 (Asn 112) places the oligosaccharide at the bottom of domain 2, where it might interact with the cell surface, opposite the highly basic surface of the GFCC' sheet.

Molecular association in the crystal

Of the contacts between sCD2 molecules in the crystal, the packing of molecules related by crystallographic 2-fold rotational symmetry is notable for the large area (650 to 690 Å² buried per molecule) and intimacy of the association (Fig. 5). There are two crystallographically independent copies of the sCD2 molecule in the crystal and it is striking that this association occurs, in similar forms, for pairs of both copy 1 and copy 2 molecules (the two versions are compared in Fig. 5a). The interaction is mediated by the domain 1 GFCC' sheet, specifically the CC' and FG loops, and as such shows some similarity to antibody V domain and CD8 dimerization, but there are also substantial differences. The usual V domain dimerization is through a specialized 'three-layer' interaction³⁵ involving side chains from the CC' and FG loop regions, primarily the C' and G strands. This interaction depends on the extensive twist introduced in these strands by conserved β -bulges. In sCD2 the equivalent β -bulges are positioned at the ends of the CC' and FG loops, thus limiting the extent of the twist, and these loop regions are no longer of similar lengths (Fig. 4b). As a result, in sCD2 the association involves main-chain hydrogen bonding between the tips of the CC' and FG loops, augmented by direct interactions between the residues on the GFCC' sheets (residues are detailed in Fig. 2). Thus the angle between the respective β -sheets in CD2 differs radically (by some 60°) from the characteristic angle of about 50° observed for V domains. In this respect, the CD2 interaction is more comparable to the standard orthogonal class of β -sheet packing.

Figure 5a compares the two independent examples of this contact by a superposition based on the first domains of molecule 1 from each pair. To superimpose the first domains of molecule 2 from each pair requires a further 18° rotation. Subtle variations between the two examples of the molecular association are accommodated without any change in the principal structural features of the interaction. The ability of V domain dimerization to exhibit a degree of structural fluidity has previously been noted³⁹.

Antibody epitopes and ligand binding sites

The epitopes for several monoclonal antibodies raised against human CD2 have been mapped to three discrete regions¹⁸.

Region 1 and 2 antibodies block ligand interactions and the combination of certain region 1 or 2 antibodies with region 3 antibodies can activate T cells. The equivalent regions in the rat sCD2 structure span residues 34–45, 77–90 and 130–131 respectively. Region 1 (the CC' loop and C' strand) and region 2 (the FG loop)^{9,16,18} are centred on opposite corners of the GFCC' sheet in domain 1. Region 3 maps to the CC' loop in domain 2 (Fig. 3a). This loop forms part of the membrane proximal region of the molecule within the highly basic environment of the domain 2 GFCC' sheet. T-cell activation or the presence of antibodies directed against the top of domain 1 is necessary to make this epitope available. But the hypothesis¹¹ that antibody binding to domain 1 could lead to conformational changes that increase the accessibility of region 3 has been argued to be very unlikely¹² and the structure now reinforces this view.

Mutagenesis of human CD2 in these regions has highlighted residues likely to be involved in binding to its ligands LFA-3 (ref. 18) and CD59 (ref. 9). The structurally equivalent residues in rat CD2 may be identified by sequence alignment¹⁶ and Fig. 5b indicates the sidechain positions for those that are exposed to solvent in the sCD2 structure. The contact site for LFA-3 involves residues in both regions 1 and 2 (ref. 18). For CD59, only the subset of residues located in region 2, the F and G strands, is implicated⁹. The LFA-3 binding site spans the diagonally opposed CC' and FG loops and thus appears to be relatively symmetrically located on the GFCC' sheet. The surface of this sheet, situated close to the top of the molecule, appears to be ideally positioned to facilitate ligand binding. This surface is walled in on either side by the CC' and FG loops. Between these walls the sheet is carpeted with a rich assortment of charged side chains, producing an electrostatically variegated but rather flat surface (Fig. 3e and f). Four of the five residues implicated in LFA-3 interaction (residues 38, 41, 81 and 85) are also intimately involved in the 2-fold lattice contact described in the previous section. This manner of association places the two C-terminal stalk regions some 140 Å apart (Fig. 5a), oriented so that their polypeptide chains can each continue into the membranes of separate cells.

Discussion

The relationship of the structure of sCD2 to its function may be illuminated by contrasting its structure with that of the NH₂-terminal fragment of CD4(D1D2). First, the element of orientational flexibility supplied by the interdomain linker region and the C-terminal stalk in sCD2 is likely to facilitate docking with molecules on opposing cell surfaces. Crystallographic data compatible with flexibility in the full extracellular region of CD4 (comprising four IgSF domains) have been reported⁴⁰. Second, sCD2 is markedly longer than CD4(D1D2), although it has fewer amino acids, so increasing the exposure of the putative ligand binding face (the domain-1 GFCC' sheet), an effect that is magnified by a third difference, the tilt of the outer domain to maintain this sheet roughly parallel to the cell surface. Flexibility and the location of ligand-binding sites close to the top of adhesion molecules may be a general feature of their organization.

The adhesion function of IgSF members is believed to predate their role in antigen recognition⁴. An evolutionary pathway for dimerization has been proposed²⁴ in which the interaction of V domains (through their GFCC' β -sheets) originated as a homophilic association having 2-fold rotational symmetry. The presence of this type of homophilic interaction as the dominant form of molecular association in the sCD2 crystals lends support to the general hypothesis. But the sCD2 association shows several key structural differences from the standard V-domain dimerization and so provides a possible paradigm for adhesion interactions. The precise position of β -bulge conformations in the C' and G β -strands and the lengths of the CC' and FG loops, combined with the relative orientations of domains 1 and 2 as determined by the linker region, have crucial long-range

effects on the geometry of dimerization, leading to a switch from a dimerization mode suitable for intermolecular adhesion to the intramolecular version seen in Fabs. Confirmation of this

adaptation of the V-domain interface in cell-adhesion interactions now awaits direct structural data for the ligand or the complex. □

Received 24 July; accepted 9 October 1992.

- Hynes, R. O. *Cell* **69**, 11–25 (1992).
- Takeichi, M. *Science* **251**, 1451–1455 (1991).
- Springer, T. A. & Lasky, L. A. *Nature* **349**, 196–197 (1991).
- Williams, A. F. & Barclay, A. N. *Rev. Immun.* **6**, 381–406 (1988).
- Dustin, M. L. & Springer, T. A. *Rev. Immun.* **9**, 27–66 (1991).
- Hunig, T. *J. exp. Med.* **162**, 890–901 (1985).
- Selvaraj, P. *et al. Nature* **326**, 400–403 (1987).
- Deckert, M., Kubar, J., Zoccola, D. & Bernard, A. *FASEB J.* **5**, A614 (abstr. 1416) (1991).
- Hahn, W. C., Menu, E., Bothwell, A. L. M., Sims, P. J. & Bierer, B. E. *Science* **256**, 1805–1807 (1992).
- Killeen, N., Moessner, R., Arvioux, J., Willis, A. & Williams, A. F. *EMBO J.* **7**, 3087–3091 (1988).
- Meier, S. C. *et al. Cell* **36**, 897–906 (1984).
- Clark, S. J., Law, D. A., Paterson, D. J., Puklavec, M. & Williams, A. F. *J. exp. Med.* **167**, 1861–1872 (1988).
- He, Q., Beyers, A. D., Barclay, A. N. & Williams, A. F. *Cell* **54**, 979–984 (1988).
- Spruyt, L. L., Glennie, M. J., Beyers, A. D. & Williams, A. F. *J. exp. Med.* **174**, 1407–1415 (1991).
- Beyers, A. D., Spruyt, L. L. & Williams, A. F. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2945–2949 (1992).
- Driscoll, P. C., Cyster, J. G., Campbell, I. D. & Williams, A. F. *Nature* **353**, 762–765 (1991).
- Rechny, M. A., Neidhardt, E. A., Sayre, P. H., Ciardelli, T. L. & Reinherz, E. L. *J. biol. Chem.* **265**, 8542–8549 (1990).
- Peterson, A. & Seed, B. *Nature* **329**, 842–846 (1987).
- Sayre, P. H., Hussey, R. E., Chang, H.-C., Ciardelli, T. L. & Reinherz, E. L. *J. exp. Med.* **169**, 995–1009 (1989).
- Matsui, K. *et al. Science* **254**, 1788–1791 (1991).
- Weber, S., Trautner, A., Oliveri, F., Gerhard, W. & Karjalainen, K. *Nature* **356**, 793–796 (1992).
- Stanton, D. E. *et al. J. exp. Med.* **169**, 1087–1099 (1989).
- Wong, Y. W., Williams, A. F., Kingsmore, S. F. & Seldin, M. F. *J. exp. Med.* **171**, 2115–2130 (1990).
- Springer, T. A. *Nature* **353**, 704–705 (1991).
- Davis, S. J. *et al. Protein Engng.* (in the press).
- McPherson, A. *Preparation and Analysis of Protein Crystals* (Wiley, New York, 1982).
- Stanley, P. *Molec. cell. Biol.* **9**, 377–383 (1989).

- Wang, B. C. *Meth. Enzym.* **115**, 90–111 (1985).
- Bricogne, G. *Acta crystallogr.* **A32**, 832–847 (1976).
- Brunker, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
- Wang, J. *et al.* **348**, 411–418 (1990).
- Ryu, S.-E. *et al.* **348**, 419–426 (1990).
- Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. *Biochemistry* **14**, 4943–4952 (1975).
- Leahy, D. J., Axel, R. & Hendrickson, W. A. *Cell* **68**, 1145–1162 (1992).
- Chothia, C., Novotny, J., Brucoleri, R. & Karplus, M. *J. molec. Biol.* **186**, 651–663 (1985).
- Fleury, S. *et al. Cell* **66**, 1037–1049 (1991).
- Saul, F. A. *et al. J. biol. Chem.* **253**, 585–597 (1978).
- Williams, A. F., Barclay, A. N., Clark, S. J., Paterson, D. J. & Willis, A. C. *J. exp. Med.* **165**, 368–380 (1987).
- Colman, P. M. *Adv. Immun.* **43**, 99–132 (1988).
- Kwong, P. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 6423–6427 (1990).
- Harlos, K. *J. appl. Crystallogr.* **25**, 536–538 (1992).
- Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. *J. molec. Biol.* **134**, 109–142 (1979).
- Jones, E. Y. & Stuart, D. I. in *Isomorphous Replacement and Anomalous Scattering* (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 39–47 (SERC Daresbury Laboratory, Warrington, 1991).
- Jones, E. Y., Walker, N. P. C. & Stuart, D. I. *Acta crystallogr.* **A47**, 753–770 (1991).
- Jones, A. *Meth. Enzym.* **115**, 157–171 (1985).
- Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
- Priestle, J. P. *J. appl. Crystallogr.* **21**, 572–576 (1988).
- Gilson, M., Sharp, K. & Honig, B. *J. comp. Chem.* **9**, 327–335 (1987).
- Yagita, H. *J. exp. Med.* (in the press).

ACKNOWLEDGEMENTS. We thank the staff at the SRS Daresbury for help with data collection; R. Bryan for computing facilities; S. Lee for assistance with the illustrations; A. N. Barclay, M. H. Brown, J. G. Cyster, C. Edge, R. Esnouf, J. Grimes, D. W. Mason and C. Somoza for discussion; and P. Driscoll, I. Campbell, who also supplied coordinates, and P. A. van der Merwe for Fig. 2. D.I.S. is a member of the Oxford Centre for Molecular Sciences supported by the SERC and MRC. E.Y.J. is supported by a Royal Society University Research Fellowship and K.H. by the OCMs. S.J.D. was funded by the UK MRC AIDS Directed Programme. Atomic coordinates have been deposited in the Protein Data Bank.

LETTERS TO NATURE

Disappearance of coronal X-ray emission in stars with cool dense winds

Bernhard Haisch^{*†}, J. H. M. M. Schmitt[†]
& A. C. Fabian[‡]

^{*}Lockheed Solar and Astrophysics Laboratory, Division 91–30, Building 252, 3251 Hanover Street, Palo Alto, California 94304, USA
[†]Max-Planck-Institut für Extraterrestrische Physik, Garching, Germany
[‡]Institute of Astronomy, Cambridge University, Cambridge CB3 0HA, UK

THE Einstein Observatory survey of cosmic X-ray sources a decade ago showed that coronae were common among diverse types of star, and were in many cases more energetic than the Sun's corona. Such coronae seemed, however, to disappear abruptly across a 'dividing line' in the Hertzsprung–Russell (H–R) diagram describing the evolution of intermediate-mass stars towards the red giant phase^{1–5}. Here we use results from the Rosat all-sky survey, which increases by an order of magnitude the number of X-ray stars, to show that the dividing line is not an artefact of poor sampling. Optical and ultraviolet observations show that the dividing line in the H–R diagram coincides approximately with the onset of cool, massive stellar winds^{6–15}, but we show that these winds are not sufficiently dense for simple X-ray absorption to be the cause of the disappearance of coronal emission. We conclude, therefore, that the dividing line represents a true evolutionary transition in these stars, at which the hot coronae are replaced by cool winds.

The concept of a dividing line in the H–R diagram was proposed to describe the apparently abrupt disappearance of coronae as stars evolve towards the giant branch at early K spectral type. As the ubiquity of stellar coronae had only just been discovered with the Einstein Observatory, the absence of coronae for stars of this type was tantalizing, especially as they seemed to disappear rather rapidly (on an evolutionary time-

TABLE 1 Hydrogen column densities in the stellar wind

Spectral type	r_0	v	M	N_H
K5 III	25	50	10^{-10}	3.4×10^{19}
K5 III	25	25	10^{-9}	6.9×10^{20}
K5 I	400	50	10^{-10}	2.2×10^{18}
K5 I	400	25	10^{-9}	4.3×10^{19}
M2 I	800	50	10^{-8}	1.1×10^{20}
M2 I	800	25	10^{-7}	2.2×10^{21}

scale). This was expected to be a clue to changes in stellar magnetic fields.

The Rosat X-ray All-Sky Survey¹⁶ has now achieved roughly the same sensitivity over the entire sky ($f_X = (1-2) \times 10^{-13}$ erg cm⁻² s⁻¹) as the typical Einstein Imaging Proportional Counter fields, which covered less than 10% of the sky. The preliminary analysis of the 70% complete survey has been discussed¹⁷. Almost 900 single GKM giants/supergiants in the Bright Star Catalogue were not detected in the X-ray survey; these were well distributed over the H–R diagram (Fig. 1, top panel). In sharp contrast, only one evolved star out of 65 detected in the 70% complete survey lay to the right of spectral type K3: this was HR4289, a little-studied K5 III showing no evidence of being a binary, whose detection involved a fairly typical 370-s cumulative effective exposure.

We have now examined the complete survey for single GKM I to IV stars, yielding 31 new detections, all of which lie to the left of the dividing line. The entire sample of 96 detections is shown in Fig. 1 (middle panel), an updated version of the 65 detections shown in ref. 17.

The possibility of a connection between the disappearance of coronae and the onset of massive winds has been discussed for many years^{1,18}. A recent compilation of stellar wind properties¹³ shows how closely the two situations may be related (Fig. 1, bottom panel). It is logical to ask whether attenuation of soft X-rays in such a wind might account for the dividing line. We

TABLE 2 Absorption of coronal emission in Rosat 0.1–2 keV band

log N_H	Total PSPC count rate		Active region	10^6 K
	10^7 K	Flare-like		
0	10.00	10.00	10.00	10.00
18	9.95	9.95	9.94	9.91
19	9.52	9.51	9.39	9.16
20	6.84	6.75	5.94	4.69
21	2.47	2.47	1.22	0.13
22	0.49	0.40	0.10	1.5×10^{-4}

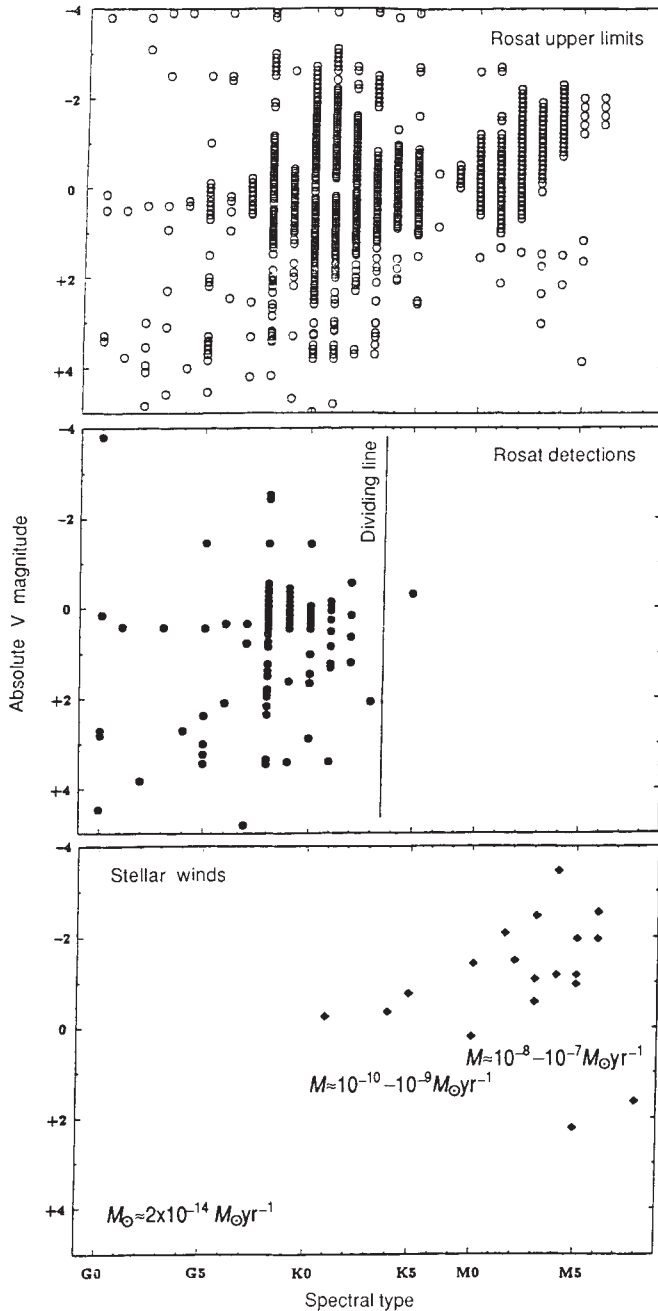


FIG. 1 H-R diagrams of late-type evolved stars. Top, Rosat All-Sky Survey non-detections of 868 non-binary stars from the Bright Star Catalogue demonstrating the observational sampling of the various spectral types. Middle, the 96 X-ray detected non-binary stars showing the cut-off of coronal emission at the dividing line. Bottom, stars having measurable stellar winds, with the general trend of mass loss in units of solar masses $M_\odot$ per year.

test this hypothesis in the most conservative way, by testing the viability of an absorption model with the maximum chance of succeeding: we assume neutral rather than partially ionized hydrogen so that absorption is maximum, and allow for very soft X-ray spectra.

The density of a spherically symmetric wind at radius r is

$$\rho(r) = \frac{\dot{M}}{4\pi r^2 v(r)} \quad (1)$$

where $\dot{M}$ is mass flow rate and $v(r)$ is the local wind velocity. The mass column density along a line-of-sight for a constant outflow at velocity v is therefore

$$\int_{r_0}^{\infty} \rho(r) dr = \frac{\dot{M}}{4\pi v} \int_{r_0}^{\infty} \frac{1}{r^2} dr = \frac{\dot{M}}{4\pi v} \frac{1}{r_0} \quad (2)$$

In units of solar masses $M_\odot \text{ yr}^{-1}$ for $\dot{M}$, solar radii $R_\odot$ for r_0 and km s^{-1} for v , and if the outflowing matter is neutral hydrogen, the column density is

$$N_H = \frac{7.2 \times 10^8}{1.66 \times 10^{-24}} \frac{\dot{M}}{v r_0} = 4.3 \times 10^{32} \frac{\dot{M}}{v r_0} \quad (3)$$

Table 1 shows the column densities, N_H , calculated from equation (3) for typical stellar parameters and mass loss rates as compiled in ref. 13. For $N_H \approx 10^{20}$, soft X-ray attenuation is generally considered to be significant. This would naturally affect the interpretation of the coronal dividing line. We investigate this quantitatively using emission models for coronal plasmas and the known response function of the Rosat Position Sensitive Proportional Counter (PSPC) telescope system.

To calculate the X-radiation from an optically thin plasma we use the line emissivities of Mewe *et al.*¹⁹ and a formula of Mewe²⁰ for the continuum emission due to free-free and free-bound charged-particle interactions, and two-photon processes. This yields the function $\Phi(T, \lambda)$ where T is temperature and λ is wavelength, which is the emissivity per unit differential emission measure, $\text{DEM} = n_e^2(T) dV$.

Stellar coronae are usually analysed and modelled as either one- or two-temperature plasmas. The lower-temperature component is typically like a solar active region; the hotter component has flare-like temperature properties. The flare-like component of a stellar corona is, of course, spectrally harder than the active-region-like component. Given the steep decrease with wavelength of the extreme ultraviolet/X-ray absorption cross-

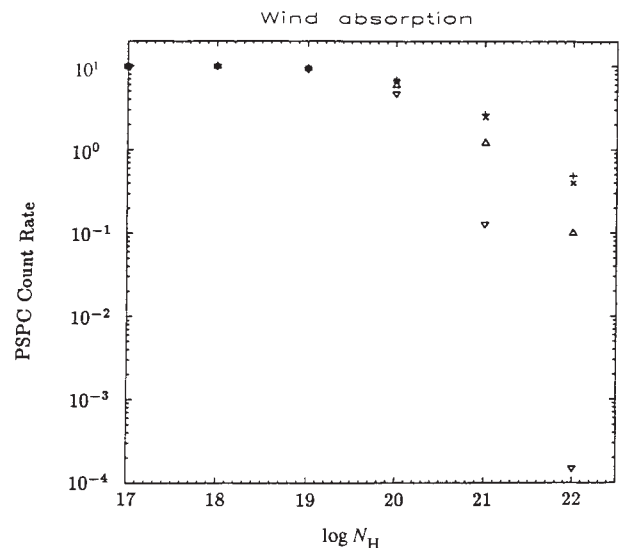


FIG. 2 Absorption of coronal X-rays in the Rosat PSPC passband for four types of spectra, 10^7 K isothermal (+), flare-like (x), active-region-like (Δ), 10^9 K (∇), as a function of N_H normalized to 10 counts per second for the unattenuated spectra.

section, σ_λ , it would take a much higher column density N_H to attenuate this flare-like component significantly. To cover this range of possibilities we concentrate on four DEMs: a 10^7 K isothermal one resulting in the hardest spectrum; a solar flare-like DEM taken from ref. 21; the solar active region DEM of ref. 22, modified at the highest temperature to decrease to zero above 10^7 K; and an isothermal DEM at 10^6 K resulting in a very soft spectrum.

For modelling the absorption, the σ_λ function adapted from Paresce in Zombeck²³ is entirely adequate, as we are interested in the overall attenuation across the fairly broad Rosat 0.1–2 keV PSPC passband, rather than in detailed modelling of absorption edges and so on. We note that the PSPC effective area has a minimum at the carbon edge at 0.28 keV. We have folded unattenuated versions of the four spectral distributions through the PSPC effective areas to generate a scale factor which, when applied to the simulation, will result in a predicted output of 10 counts per second. Next, we attenuate each of the four spectra, normalized by the scale factor, by five different N_H column densities. The results are listed in Table 2 and shown in Fig. 2.

In the case of the harder spectra, even high column densities ($N_H = 10^{22}$) would only attenuate the coronal emission by a factor of 20–25. This column density could cause factor of 100 attenuation in an active-region-like spectrum. Even for the coolest corona (10^6) with the softest spectrum, $N_H > 10^{21}$ would be required to attenuate the X-ray luminosity L_X by a factor of 100.

In the case of Arcturus, the measured Rosat upper limit (achieved during a 18.6-ks targeted observation²⁴) is $L_X < 3 \times 10^{25}$ erg s⁻¹. This corresponds to a ratio $\log f_X/f_V < -9.15$, where f_X is X-ray flux at Earth and f_V is V-band flux at Earth. As discussed also in ref. 17, Arcturus is the archetypal old population II non-coronal red giant. On the other hand there are early K coronal giant stars in the Rosat survey and in the Einstein IPC fields²⁵ having $-4.2 > f_X/f_V > -6.2$. If Arcturus did possess such a corona, then the Rosat upper limit would imply attenuation in the range of 10^2 – 10^5 , requiring $N_H \approx 10^{21}$ – 10^{23} cm⁻². As is evident from Table 1 and Fig. 1, this is considerably more mass loss than is thought to occur for early K giants.

What possibilities remain open? Might red giants (to the right of the dividing line) be on average more distant than yellow giants (left of the dividing line)? This is unlikely, as the absolute magnitude of these stars does not vary much across the dividing line⁵, and in any case would not affect the f_X/f_V ratio. Another possibility that we cannot categorically rule out is that an accelerating wind would have a higher column density than the constant velocity winds modelled here. Little is known about the actual velocity profiles of winds near the surfaces of stars, so we cannot pursue this further.

What physical processes might explain the dividing line? On theoretical grounds, it has been proposed²⁶ that hot coronal loops become unstable in evolved stars because of the combined effects of large size (presumed to scale as the stellar radius R_*) and low surface gravity, leaving only low-lying cool loops which do not emit X-rays. But the growing body of evidence that the onset of massive winds occurs right up to the dividing line suggests that a phase transition is occurring from a hot, magnetically confined corona to a cool, open, massive wind outflow. This was in fact the original explanation proposed by Linsky and Haisch¹. Subsequent attempts to model this in terms of winds driven by Alfvén waves met with limited success¹⁸, but interest is now developing in a theoretical approach involving reflection of Alfvén waves²⁷. The existence of a coronal dividing line remains a challenge to our understanding of dynamo mechanisms and coronal heating. □

4. Simon, T., Linsky, J. L. & Stencel, R. E. *Astrophys. J.* **257**, 225–246 (1982).
5. Haisch, B. in *Lecture Notes in Physics*, Vol. 291, 269–282 (1987).
6. Reimers, D. & Schmitt, J.H.M.M. *Astrophys. J.* **392**, L55–L58 (1992).
7. Reimers, D. *Astr. Astrophys.* **57**, 395–400 (1977).
8. Hartmann, L., Dupree, A. K. & Raymond, J. C. *Astrophys. J.* **236**, L143–L147 (1980).
9. Judge, P. G., Jordan, C. & Rowan-Robinson, M. *Mon. Not. R. astr. Soc.* **24**, 93–106 (1987).
10. Brown, A., Drake, S. A., van Steenberg, M. E. & Linsky, J. L. *Astrophys. J.* **373**, 614–619 (1991).
11. Dupree, A. K. & A. Rev. *Astr. Astrophys.* **24**, 377–420 (1986).
12. Lafon, J.-P. J. & Berruyer, N. *Astr. Astrophys. Rev.* **2**, 249–289 (1991).
13. Judge, P. G. & Stencel, R. E. *Astrophys. J.* **371**, 357–379 (1991).
14. Hartmann, L., Jordan, C., Brown, A. & Dupree, A. K. *Astrophys. J.* **296**, 576–592 (1985).
15. Haisch, B., Schmitt, J. H. M. M. & Rosso, C. *Astrophys. J.* **388**, L61–L64 (1992).
16. Trümper, J. et al. *Nature* **349**, 579–583 (1991).
17. Haisch, B., Schmitt, J. H. M. M. & Rosso, C. *Astrophys. J.* **383**, L15–L18 (1991).
18. Haisch, B., Linsky, J. L. & Basri, G. S. *Astrophys. J.* **235**, 519–533 (1980).
19. Mewe, R., Gronenschild, E. H. B. M. & van den Oord, G. H. J. *Astr. Astrophys.* **62**, 197–254 (1985).
20. Mewe, R. *Astr. Astrophys. Rev.* **3**, 127–168 (1991).
21. Bruner, M. E. & McWhirter, R. W. P. *Astrophys. J.* **326**, 1002–1016 (1988).
22. Dere, K. P. *Solar Phys.* **77**, 77–93 (1982).
23. Zombeck, M. *Handbook of Space Astronomy and Astrophysics*, 110 (Cambridge Univ. Press, 1982).
24. Ayres, T. R., Fleming, T. A. & Schmitt, J. H. M. M. *Astrophys. J.* **376**, L45–L48 (1991).
25. Maggio, A. et al. *Astrophys. J.* **348**, 253–278 (1990).
26. Antiochos, S. K., Haisch, B. M. & Stern, R. A. *Astrophys. J.* **307**, L55–L59 (1986).
27. Rosner, R., An, C.-H., Musielak, Z. E., Moore, R. L. & Suess, S. T. *Astrophys. J.* **372**, L91–L94 (1991).

ACKNOWLEDGEMENTS. B.H. thanks J. Trümper and the Max-Planck-Institut für Extraterrestrische Physik for hospitality.

Weakly damped Alfvén waves as drivers of solar chromospheric spicules

Gerhard Haerendel

Max-Planck-Institut für Extraterrestrische Physik, 8046 Garching, Germany

THE solar chromosphere, which separates the photosphere (temperature $T \approx 6,400$ K) from the hotter solar corona ($T \approx 10^6$ K), has a very inhomogeneous structure which is strongly influenced by the magnetic field. In the lowest 2,000 km of chromospheric altitude, the density falls by six orders of magnitude, and the temperature stays below 8,000 K. Above this altitude the transition to the corona is extremely irregular. It is dominated by spicules: thin (<1,000 km) protrusions of cool chromospheric material which extend, with speeds of 25 km s⁻¹ and for durations of 5–10 min, up to 10,000 km into the corona with little change in density. Their origin is not yet understood, and I suggest here that the force that propels them against gravity may be the transfer of momentum from upward-moving Alfvén waves. In the upper chromosphere the ionized plasma component is collisionally coupled to the neutral gas, but the coupling is not perfect, so that the neutral material can acquire a net velocity with respect to the ionized component. This process is known to damp the Alfvén waves¹², but I show that it can also, as the chromosphere peters out, transfer enough momentum to local volumes to create and drive the spicules.

Spicules are obviously not in simple hydrostatic equilibrium, otherwise they could only reach heights of a few pressure scale-heights (~600 km). They could, however, be jets accelerated at their base and flying up ballistically along the magnetic field lines. Several mechanisms have been proposed which, in principle, could give chromospheric matter the necessary initial kick, for example shock-wave heating¹, diamagnetic squeezing, the so-called melon-seed effect^{2,3}, overheating and thermal expansion⁴, Laval nozzle acceleration in the diverging fields of the supergranular network⁵, and magnetic reconnection⁶. But ballistic flight to heights of 10,000 km requires initial velocities of ~70 km s⁻¹, which have never been observed. Some combination of the above processes occurring over the extent of the spicules is conceivable—one example is the rebound shock model⁷—but such models tend to imply higher temperatures than observed, or differences in shape, such as a noticeable widening of the spicule diameters. The spicules seem, however,

Received 22 July; accepted 18 September 1992.

1. Linsky, J. L. & Haisch, B. *Astrophys. J.* **229**, L27–L32 (1979).
2. Ayres, T. R. et al. *Astrophys. J.* **250**, 293–299 (1981).
3. Haisch, B. & Simon, T. *Astrophys. J.* **273**, 252–258 (1982).

to be embedded in the fairly homogeneous fields of the corona, well above the level of mutual adjustment of the strongly inhomogeneous photospheric fields. The schematic illustration shown in Fig. 1 (from ref. 8) describes the spicules fairly well. The observations indicate a force acting against gravity without the intermediary of a pressure increase. This is regarded as one of the unsolved problems of solar physics⁹⁻¹¹. Here I draw attention to a physical process that yields such a force and could offer a solution to the problem.

The process under consideration is ion-neutral collisional damping of Alfvén waves. It is known for its heating capability, but to my knowledge has not been discussed for the net force that it can exert. Alfvén waves are elastic transverse waves in a magnetized plasma for which the magnetic tension delivers the restoring force and the plasma density, ρ , the inertia. They travel along the magnetic field, $\mathbf{B}$, with a characteristic speed, the Alfvén speed, $v_A = B/\sqrt{4\pi\rho}$. In the chromosphere such waves find special conditions, because the medium is only partially ionized; the collisional coupling between neutrals and ions is so strong that the sum of ions and neutrals provides the inertial force, and $\rho = \rho_i + \rho_n$ has to be inserted in the expression for v_A . This applies as long as the wave frequency, ω , is small compared with the neutral-ion collision frequency, ν_{ni} . Whatever motion is imparted to the ions by the magnetic forces, the neutrals will follow almost immediately. The essential word here is 'almost'. Collisional coupling takes a finite time; hence there is a small slippage between neutrals and ions which is proportional to the ratio ω/ν_{ni} .

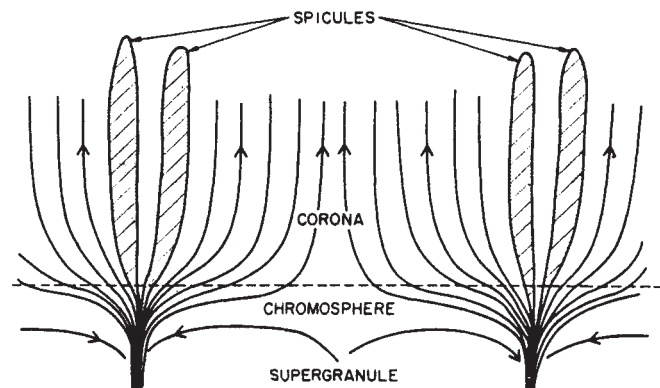
A finite velocity difference between ions and neutrals allows momentum and energy transfer, and thus causes wave damping. (The same principle underlies many technical devices such as nutation dampers.) The accompanying heat deposition, although very small per wavelength or wave period, can be considerable; Piddington¹² proposed in 1956 that it was important for the chromosphere. This process will dominate Alfvén wave damping in the upper chromosphere¹³ and may be the main heat source there. In such an environment the phase velocity, $v_{ph} = \omega/k$, of Alfvén waves must be slightly modified^{12,14,15}

$$v_{ph} = \frac{\omega}{k} = v_A \left(1 + i \frac{\rho_n}{\rho_i + \rho_n} \frac{\omega}{\nu_{ni}} \right)^{1/2} \quad (1)$$

The imaginary part in the brackets constitutes the damping term at a rate

$$\gamma = \frac{1}{2} \frac{\rho_n}{\rho_i + \rho_n} \frac{\omega^2}{\nu_{ni}} \quad (2)$$

This implies that a propagating wave will be damped over a length, $\lambda_{damp} = v_A/\gamma$. As I assume throughout that $\omega \ll \nu_{ni}$, which for the upper chromosphere means that I consider waves with frequencies well below 10 Hz, λ_{damp} is very large compared with a wavelength v_A/ω .



To demonstrate the property of interest, we have to consider the electric current, $\mathbf{j}$, connected with the wave field. I take a homogeneous magnetic field, $\mathbf{B}$, in the z -direction and a plane wave with velocity vector, v_x , and magnetic perturbation, B_x , in the x -direction. In a plasma in which the ions are strongly magnetized (that is, where they perform many Larmor rotations before colliding), their transverse velocity, v_x , is equivalent to the existence of an electric perturbation field, E_y

$$E_y = \frac{v_x}{c} B_z \quad (3)$$

($B_z \approx |\mathbf{B}| = B$, $B_x \ll B$). Current and electric field for Alfvén waves in partially ionized plasmas with $\omega \ll \nu_{ni}$ are related by

$$j_y = \frac{(\rho_i + \rho_n)c^2}{B^2} \left(i\omega + \frac{\rho_n}{\rho_i + \rho_n} \frac{\omega^2}{\nu_{ni}} \right) E_y \quad (4)$$

The first term in the brackets ($i\omega$) stems from the inertia of the mass oscillating in the wave field. Current and electric field (or velocity v_x) are 90° out of phase. The second term, however, contains the effect of momentum coupling to the neutrals which depends on the slippage (ω/ν_{ni}). It is in phase with E_y as in an ohmic resistor. If one adds Faraday's law, which yields a relation between the magnetic perturbation field, B_x , and E_y

$$B_x = -\frac{c}{v_{ph}} E_y \quad (5)$$

and averages over many wave periods, one obtains a residual net force

$$\bar{F}_z = -\frac{1}{c} \overline{j_y B_x} = \rho_n \overline{v_x^2} \frac{\omega^2}{v_A \nu_{ni}} \quad (6)$$

$\bar{F}_z$ carries the sign of the phase velocity. It is proportional to the neutral density and the mean square of the wave's velocity (or electric field) amplitude, v_x . The accompanying energy dissipation rate is:

$$\epsilon_w = \overline{j_y E_y} = \rho_n \overline{v_x^2} \frac{\omega^2}{\nu_{ni}} = v_A \bar{F}_z \quad (7)$$

The inertial part of the current (the $i\omega$ term in equation (4)), which is non-dissipative, constitutes a reversible interchange of momentum and energy between magnetic field and plasma (both ions and neutrals). It vanishes in the averaging process. Collisions and a finite slippage are needed to obtain a net force along the wave propagation vector. Dissipation and force are simply related by the wave's phase velocity.

I now turn to the situation in the chromosphere. Can the net force, $\bar{F}_z$, become comparable to the gravitational force, $\rho g_\odot$ ($g_\odot = 2.7 \times 10^4 \text{ cm s}^{-2}$)? We adopt the following parameters for a typical spicule: $\mathbf{B} = 10 \text{ G}$, number density of ions $n_i = 3.5 \times 10^{10} \text{ cm}^{-3}$, neutral density $n_n = 1.5 \times 10^{10} \text{ cm}^{-3}$ (ref. 16), and a mix of H I, H II and He I taken from Vernazza *et al.*¹⁷. With collision cross-sections for H-H⁺ of $4 \times 10^{-15} \text{ cm}^2$ and H⁺-He of $8 \times 10^{-16} \text{ cm}^2$ (ref. 13), I calculate $\nu_{ni} \approx 70 \text{ s}^{-1}$. $\bar{F}_z = (\rho_i + \rho_n)g_\odot$

FIG. 1 Schematic illustration of the magnetic field concentrated at the boundaries of supergranules and expanding into the corona with spicules embedded (after ref. 8).

and equation (6) imply:

$$\frac{\overline{v_x^2}}{\rho_n} = \frac{\rho_i + \rho_n}{\rho_n} \frac{v_A \nu_{ni} g_\odot}{\omega^2} = \left(\frac{12.6 \text{ km s}^{-1}}{f_{\text{Hz}}} \right)^2 \quad (8)$$

where f_{Hz} is the frequency $\omega/2\pi$ measured in hertz. The Alfvén velocity, v_A , is 100 km s^{-1} for the above parameters. For the wave amplitude, $\sqrt{\overline{v_x^2}}$, to stay below v_A , f must exceed 0.1 Hz. But f should not be too great, because the damping rate increases in proportion to f^2 (equation (2)). To drive chromospheric material up the field lines by 10,000 km or so, λ_{damp} should not fall below this length. This sets an upper limit to ω

$$\lambda_{\text{damp}} = \frac{v_A}{\gamma} = \frac{2(\rho_i + \rho_n)}{\rho_n} \frac{\nu_{ni} v_A}{\omega^2} \geq 10^4 \text{ km} \quad (9)$$

which implies $f \leq 0.35 \text{ Hz}$.

Alfvén wave damping can thus drive the chromospheric spicules if high-amplitude waves ($v_x/v_A \approx 0.3$) exist at frequencies of $\sim 0.3 \text{ Hz}$. Of course, my numbers are not precise, but they are sufficiently realistic. One can even find observational support in measured Doppler broadenings of individual spicules¹⁶ showing $\delta\lambda/\lambda \approx 10^{-4}$, corresponding to $\pm 30 \text{ km s}^{-1}$.

A final consistency check comes from the energy balance. The dissipated energy, $\dot{\epsilon}_w$, is easily estimated with help of equation (7). The force, $\vec{F}_z$, exerted by the Alfvénic wave drag is mainly needed to balance the gravitational force. Hence $\dot{\epsilon}_w \approx v_A \rho g_\odot = 2.2 \times 10^{-2} \text{ erg cm}^{-3} \text{ s}^{-1}$. Cooling is mostly radiative, as adiabatic expansion can contribute only a few per cent. For a radiative cooling rate of the form $\dot{\epsilon}_R = n_e n_H Q(T)$, where n_e is the electron density ($n_e = n_i$) and n_H the total hydrogen density¹⁸, the energy balance can be expressed by

$$v_A \rho g_\odot \approx n_e n_H Q(T) \quad (10)$$

We solve for the temperature, T , by using the approximation $Q(T) = \chi T^{6.15}$, which is valid between 8,000 K and 20,000 K (ref. 19)

$$T_{\text{sp}} \approx 13,500 \text{ K} \left(\frac{3.5 \times 10^{10} \text{ cm}^{-3} \times v_A}{n_e \times 100 \text{ km s}^{-1}} \right)^{0.163} \quad (11)$$

The temperature values implied by equation (11) are in good agreement with observations and carefully derived thermodynamic models¹⁶.

Weak damping of intense Alfvén waves by ion-neutral collisions can therefore create a drag force capable of lifting cool, dense matter high up into the corona. Not only does this process promise to solve the riddle of spicules, but it may also be essential for other phenomena in which cool and relatively dense matter is sustained inside the hot and dilute corona. Some questions remain, the most important of which concern the origin of the Alfvén waves (see ref. 13) and their further effects as they propagate, only partly damped, out into the corona. $\square$

Direct measurement of potassium-promoted change in heat of adsorption of CO on Ni{100}

N. Al-Sarraf, J. T. Stuckless & D. A. King

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

THE action of alkali metals and their oxides as promoters for many catalytic reactions has been known for nearly a century¹. Alkali promotion of carbon monoxide reactivity, for example, forms part of the industrially important synthesis of hydrocarbons, and is thought to follow from a weakening of the internal C–O bond resulting from an enhanced binding interaction between the CO molecule and the catalyst². Several spectroscopic techniques have been used to examine this effect for CO adsorption on single-crystal metal surfaces^{2–6}, but bond energy changes can only be inferred indirectly from these techniques. Here we report direct measurements, using a recently designed single-crystal adsorption microcalorimeter^{7,8}, of the change in CO adsorption heat on a Ni{100} surface for a range of potassium precoverages. We find that the effect of promotion is unexpectedly large, with the adsorption heat increasing from a clean-surface value of 124 kJ mol^{-1} to $\sim 310 \text{ kJ mol}^{-1}$ at high potassium precoverages. We attribute this effect to a combination of stronger CO binding to the surface and increased ionization of the potassium adatoms.

The ultra-high-vacuum single-crystal adsorption microcalorimeter^{7,8} has three principal features. (1) Gas dosing is achieved using a pulsed molecular beam source, with 50-ms pulses, each containing $\sim 10^{12}$ molecules, equivalent to $\sim 2\%$ of a molecular monolayer. The dosage and sticking coefficient, and hence the coverage increment, is measured for each pulse. (2) The Ni{100} single crystal, 6 mm in diameter, is only $\sim 2,000 \text{ Å}$ thick; it is mounted on a nickel ring with a 4-mm hole in its centre. The heat capacity of the central 2-mm-diameter section intersected by the collimated molecular beam is only $\sim 1 \mu\text{J K}^{-1}$. (3) The temperature change in the crystal during adsorption of a gas pulse is measured remotely by focusing infrared radiation from the crystal back face onto a sensitive a.c. coupled detector. Calibration is done *in situ* using a chopped laser beam to provide a known heat increment. Measurements of CO adsorption on the bare nickel surface, for which the adsorption heat is well known⁹, indicate that the calorimeter is accurate to within its precision limit of $\pm 5 \text{ kJ mol}^{-1}$.

Crystal cleanliness and crystallinity were checked by Auger electron spectroscopy and low energy electron diffraction, respectively. Potassium was dosed onto the crystal using a getter source, and coverages were obtained by comparing the Auger peak ratios of K and Ni. All experiments were done with the crystal at room temperature.

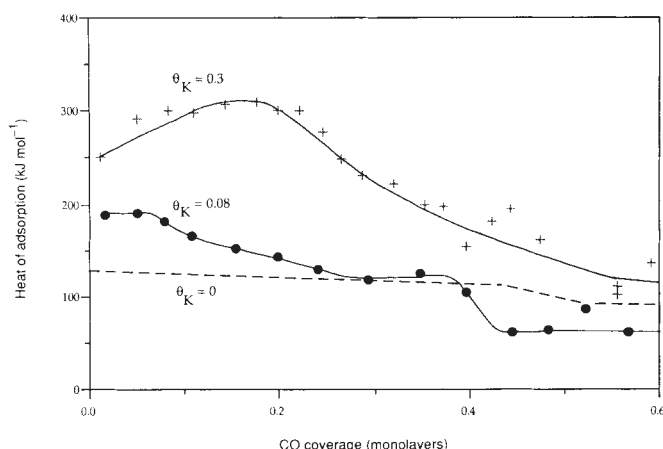
The results obtained fall into two categories, depending on the potassium precoverage. For alkali adsorption on transition metals, the work function initially falls rapidly with increasing alkali metal coverage, then passes through a minimum, and finally increases to that of the alkali metal itself. This behaviour is attributed to an initial charge transfer from the alkali adatom to the transition metal, which is then reversed at high coverages as the adlayer changes from cationic to covalent or metallic behaviour¹⁰. The two categories of CO adsorption heats are based on the K precoverage being higher or lower than that at the work function minimum.

Figure 1 shows the results obtained for the initially clean Ni substrate, and for low and high K precoverages. When the K precoverage is 0.08 monolayers, the initial heat of adsorption is 190 kJ mol^{-1} , which is significantly higher than the clean

Received 24 August; accepted 6 October 1992.

1. Parker, E. N. *Astrophys. J.* **140**, 1170–1173 (1964).
2. Schlüter, A. *IAU Symp. No. 4*, Radio Astronomy (ed. van de Hulst, H. C.) 356–357 (Cambridge Univ. Press, 1957).
3. Parker, E. N. *Astrophys. J. Suppl.* **27**, 51–76 (1957).
4. Bessey, R. J. & Kuperus, M. *Solar Phys.* **12**, 216–228 (1970).
5. Uchida, Z. *Publ. astr. Soc. Japan* **21**, 218–240 (1969).
6. Pikelner, S. B. *Astr. Zh.* **46**, 328–336; *Sov. Astr. Astrophys.* **13**, 259–266 (1969).
7. Sterling, A. C. & Hollweg, J. V. *Astrophys. J.* **327**, 950–963 (1988).
8. Kopp, R. A. & Kuperus, M. *Solar Phys.* **4**, 212–223 (1968).
9. Priest, E. *Solar Magnetohydrodynamics* (Reidel, Dordrecht, 1982).
10. Athay, R. G. in *Physics of the Sun Vol. II* (ed. Sturrock, P. A. et al.) 51–69 (Reidel, Dordrecht, 1986).
11. Foukal, P. V. *Solar Astrophysics* (Wiley, New York, 1990).
12. Piddington, J. H. *Mon. Not. R. astr. Soc.* **116**, 314–323 (1956).
13. Osterbrock, D. E. *Astrophys. J.* **134**, 347–388 (1961).
14. Dungey, J. W. *Cosmic Electrodynamics* (Cambridge Univ. Press, 1958).
15. Hasegawa, A. & Uberoi, C. *The Alfvén Wave* (DOE, Springfield, Virginia, 1982).
16. Beckers, J. M. A. *Rev. Astr. Astrophys.* **10**, 73–100 (1972).
17. Vernazza, J. E., Avrett, E. H. & Loesser, R. *Astrophys. J. Suppl.* **45**, 635–725 (1981).
18. McWhirter, R. W. P., Thonemann, P. C. & Wilson, R. *Astr. Astrophys.* **40**, 63–73 (1975).
19. Peres, G., Rosner, R., Serio, S. & Vaiana, G. S. *Astrophys. J.* **252**, 791–799 (1982).

ACKNOWLEDGEMENTS. We thank R. Treumann and A. Krüger for discussions.



surface value of 124 kJ mol^{-1} . It remains constant up to a CO coverage of 0.05 monolayers before starting to fall, eventually returning to the clean surface value at a fractional coverage $\theta_{\text{CO}} = 0.35$. It then falls again, stabilizing at a mean value of 70 kJ mol^{-1} , corresponding to reversible CO chemisorption, which occurs at room temperature once steady-state coverage is attained.

The CO adsorption heat at higher potassium coverages reveals very different behaviour, as is seen in Fig. 1 for $\theta_K = 0.3$ monolayers. The initial heat of adsorption is considerably higher, at 250 kJ mol^{-1} , and as the CO coverage increases the adsorption heat increases further, reaching a maximum of 310 kJ mol^{-1} at a CO coverage of 0.16 monolayers. Following this peak, the heat falls steadily to $\theta_{\text{CO}} \approx 0.5$ monolayers before levelling off as before. Figure 2 summarizes the overall trend in both the initial and the maximum heats of CO adsorption for several potassium precoverages.

The increase of 66 kJ mol^{-1} in the initial CO heat of adsorption for low K coverages correlates well with the reported shift of the CO thermal desorption peak from $\sim 450 \text{ K}$ without K to $\sim 700 \text{ K}$ with pre-adsorbed K (ref. 11). It has, however, been suggested that a similar shift in desorption temperature, after alkali promotion of a Pt{111} surface, is due mainly to a decrease in the desorption frequency factor rather than to an increased activation energy¹². Our results show directly that the heat of adsorption is significantly increased. It should be noted that at low K coverages CO is known to bind preferentially to the promoted sites⁴, so that the heat released at low CO coverages corresponds to adsorption at K-promoted sites, and at high CO coverages to adsorption onto Ni sites far removed from K. There is sufficient mobility to produce the most stable adlayer during the adsorption pulse.

For the high K coverage regime, our results show a very large increase in the heat of adsorption, to $\sim 300 \text{ kJ mol}^{-1}$, although the simultaneous desorption of K and CO still occurs at 700 K (ref. 11). With such large adsorption heats indicating a very strong interaction, it becomes an important consideration whether the C–O bond is sufficiently weakened to produce dissociative adsorption. Spectroscopic evidence for similar surfaces indicates, however, that at temperatures below 400 K isotopic mixing of CO does not occur, implying that CO is undissociated in the presence of K (ref. 13). An explanation for the large adsorption heats which does not invoke CO dissociation is discussed below.

At low K coverages charge transfer to the transition metal leaves an ionic species on the surface^{14,15}, which can be theoretically mimicked by an external positive point charge¹⁶. CO molecules adsorbed in the vicinity of this charge will, to a first approximation, have their orbital binding energies rigidly downward-shifted, allowing increased electron transfer from the sub-

FIG. 1 Calorimetric heats of adsorption, given as the negative value of the enthalpy change, as a function of CO coverage on Ni{100}. The dashed line represents the heat of adsorption of CO on the clean surface; ● shows the heat of adsorption at a low potassium coverage, $\theta_K = 0.08$; and + shows the heat of adsorption at a high potassium coverage, $\theta_K = 0.3$.

strate into the lifetime-broadened CO $2\pi^*$ antibonding orbital¹⁷. The CO heat of adsorption is thus raised because of strengthened metal–CO bonding, although the C–O bond will, of course, be weakened. We ascribe the increase in the adsorption heat of CO from 124 to 190 kJ mol^{-1} for low θ_K to this effect, a phenomenon of direct relevance to the reactivity of CO in catalytic processes.

For higher precoverages of K, the adatom bonding is switched to essentially covalent character, with a large decrease in the K adsorption energy as the contribution due to the electrostatic interaction of the point charge with the nickel substrate is lost. For Na on Ni{100} this decrease was estimated, from experimental desorption rates, as $\sim 85 \text{ kJ mol}^{-1}$ (ref. 18). We propose that adsorption of CO into this essentially covalently bonded K adlayer screens the K adatoms from each other and brings about their re-ionization, hence increasing the K-surface bond energy, as schematically shown in Fig. 3a. There will be a significant long-range electrostatic contribution as a charged K–CO lattice is formed (Fig. 3b). This lattice, composed of the cations and anions in the adlayer and their effective image charges in the transition metal substrate, will have a Madelung energy associated with it. For a K precoverage of 0.2 monolayers, the initial 2% dose of CO is adsorbed with a heat of adsorption of 310 kJ mol^{-1} (Fig. 2). Even for this small CO coverage, mobility in the adlayer is sufficient to generate relatively large

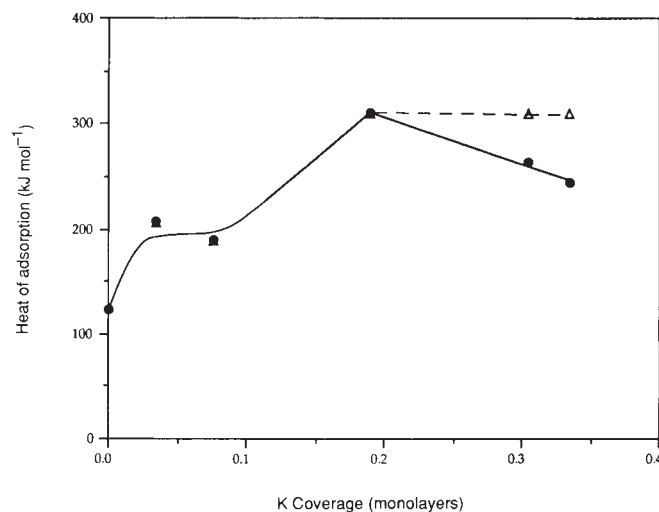


FIG. 2 Heat of adsorption results for CO adsorption as a function of increasing potassium coverage, θ_K . Solid lines (●) represent initial heats of adsorption; dashed lines (Δ) represent maximum values.

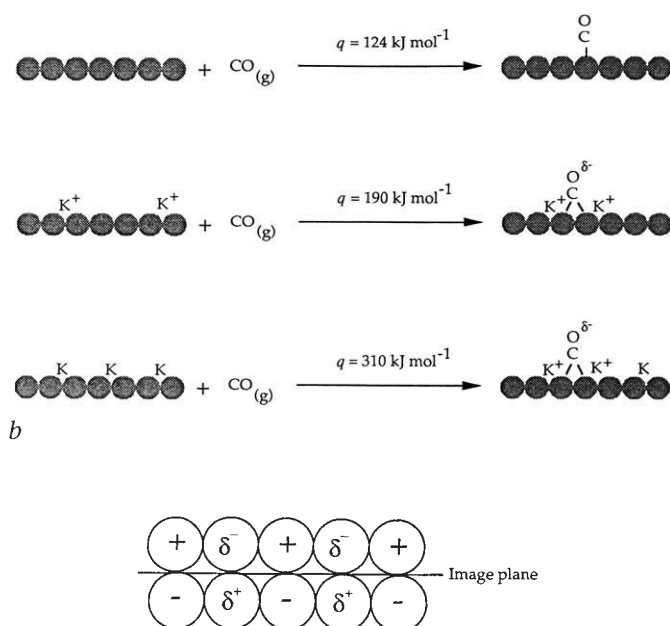


FIG. 3 a, Schematic of the CO interaction as θ_K is increased, and of the related heat of adsorption q . b, Image charge ionic bilayer lattice formed at high θ_K .

K^+CO^- islands within the 50-ms pulse, and we associate the heat released with the re-ionization of covalently bonded K adatoms and the associated electrostatic interaction terms. For larger precoverages of K, however, the heat release is smaller when CO is initially dosed onto the surface, as shown in Fig. 1 for $\theta_K = 0.3$ monolayers, and rises with increasing CO coverage to the same maximum value of 310 kJ mol^{-1} . At high K coverages there are relatively few unblocked sites on the surface, and this may reduce the mobility of adsorbed CO across the surface. Initially, therefore, at relatively high θ_K it may be that small $K^+CO^-K^+$ islands are formed. Only as these islands coalesce is the full Madelung energy involved, producing the observed increase in adsorption heat with coverage. At high K coverages there are, therefore, two dominant contributions to the CO adsorption heat: the increased CO $2\pi^*$ population¹⁷, which contributes, from our low θ_K data, a 70 kJ mol^{-1} increase; and ionization of K adatoms, contributing a further $\sim 125 \text{ kJ mol}^{-1}$ when large islands are formed. This second component, consisting of K re-ionization and Madelung terms, is important for understanding catalytic processes as it indicates that the local environment about the CO is ionic, even at high precoverages of K. $\square$

15. Benesh, G. A. & King, D. A. *Chem. Phys. Lett.* **191**, 315–319 (1992).
16. Finnis, M. W. *Surf. Sci.* **241**, 61–72 (1991).
17. Nørskov, J. K., Holloway, S. & Lang, N. D. *J. Vac. Sci. Technol. A* **3**, 1668–1672 (1985); *Surf. Sci.* **137**, 65–78 (1984).
18. Gerlach, R. L. & Rhodin, T. N. *Surf. Sci.* **19**, 403–426 (1970).

ACKNOWLEDGEMENTS. We thank J. Chevallier for thin-film single crystals. The SERC (UK) is acknowledged for an equipment grant, and for a studentship to N.A.-S.; and the NSERC (Canada) is acknowledged for a fellowship to J.T.S.

Evidence for massive discharges of icebergs into the North Atlantic ocean during the last glacial period

Gerard Bond*, Hartmut Heinrich†, Wallace Broecker*, Laurent Labeyrie‡, Jerry McManus*, John Andrews§, Sylvain Huon||, Ruediger Jantschik¶, Silke Clasen#, Christine Simet**, Kathy Tedesco‡, Mieczyslaw Klas*, Georges Bonani†† & Susan Ivy

* Lamont-Doherty Geological Observatory, Palisades, New York 10964, USA

† Bundesamt für Seeschifffahrt und Hydrographie, Postfach 30 12 20, 2000 Hamburg 36, Germany

‡ C.F.R. Laboratoire mixte CNRS-CEA, Domaine du CNRS, 91198, Gif-sur-Yvette Cedex, France

§ Institute of Arctic and Alpine Research and Geological Sciences Box 450, University of Colorado, Boulder, Colorado 80309 USA

|| Département de Minéralogie, 13 rue des Maraichers, CH-1211 Genève 4, Switzerland

¶ Institut de Géologie, 11, rue E. Argand, CH-2007 Neuchâtel, Switzerland

Abteilung Sediment-Geologie, Institut für Geologie und Paläontologie, Goldschmidtstrasse 3, W-3400 Göttingen, Germany

** Institut für Geologie und Paläontologie, Sigwartstrasse 10, 7400 Tübingen, Germany

†† Institut für Mittelenergiephysik, ETH Honggerberg, CH-8093 Zurich, Switzerland

SEDIMENTS in the North Atlantic ocean contain a series of layers that are rich in ice-rafted debris and unusually poor in foraminifera¹. Here we present evidence that the most recent six of these 'Heinrich layers', deposited between 14,000 and 70,000 years ago, record marked decreases in sea surface temperature and salinity, decreases in the flux of planktonic foraminifera to the sediments, and short-lived, massive discharges of icebergs originating in eastern Canada. The path of the icebergs, clearly marked by the presence of ice-rafted detrital carbonate, can be traced for more than 3,000 km—a remarkable distance, attesting to extreme cooling of surface waters and enormous amounts of drifting ice. The cause of these extreme events is puzzling. They may reflect repeated rapid advances of the Laurentide ice sheet, perhaps associated with reductions in air temperatures, yet temperature records from Greenland ice cores appear to exhibit only a weak corresponding signal. Moreover, the 5–10,000-yr intervals between the events are inconsistent with Milankovitch orbital periodicities, raising the question of what the ultimate cause of the postulated cooling may have been.

The most detailed study of the sediment composing the Heinrich deposits has been made on cores from DSDP site 609² (Fig. 1). High percentages of ice-rafted detritus (IRD) and low concentrations of foraminifera clearly define the six deposits that formed during the last glacial period (Fig. 2). Foraminiferal stratigraphy indicates that these deposits correlate directly with the six deposits originally identified by Heinrich in the Dreizack seamount area² (Fig. 1). We have now traced Heinrich deposits in other cores from the North Atlantic and Labrador Sea (Fig. 1), demonstrating that they are widespread features of the North Atlantic ocean. Radiocarbon measurements by

Received 26 June 1992; accepted 14 October 1992.

1. Moss, W. D. *Catal. Rev. Sci. Engng* **25**, 591–637 (1983).
2. Bonzel, H. P. *Surf. Sci. Rep.* **8** (1988).
3. Kiskinova, M. P. *Poisoning and Promotion in Catalysis based on Surface Science Concepts* (Elsevier, Amsterdam, 1992).
4. Heskett, D. *Surf. Sci.* **199**, 67–86 (1988).
5. Bonzel, H. P., Bradshaw, A. M. & Ertl, G. (eds) *Physics and Chemistry of Alkali Metal Adsorption* (Elsevier, Amsterdam, 1989).
6. Lackey, D. & King, D. A. *J. chem. Soc. Faraday I* **83**, 2001–2013 (1987).
7. Borroni-Bird, C. E. & King, D. A. *Rev. Sci. Instrum.* **62**, 2177–2185 (1991).
8. Borroni-Bird, C. E., Al-Sarraf, N., Andersson, S. & King, D. A. *Chem. Phys. Lett.* **183**, 516–520 (1991).
9. Tracy, J. C. *J. chem. Phys.* **56**, 2736–2747 (1972).
10. Gurney, R. W. *Phys. Rev.* **47**, 479–482 (1935).
11. Luftman, H. S., Sun, Y.-M. & White, J. M. *Appl. Surf. Sci.* **19**, 59–72 (1984).
12. Whitman, L. J. & Ho, W. *J. chem. Phys.* **90**, 6018–6025 (1989).
13. Kumar, S., Sands, W. D., Yates, J. T. & Janda, K. C. *J. Am. chem. Soc.* **113**, 3684–3688 (1991).
14. Scheffler, M. et al. *Physica B* **172**, 143–153 (1991).

accelerator mass spectrometry (AMS) at DSDP site 609 date the first three deposits at ~14,300, ~21,000 and ~28,000 years (Table 1), and extrapolation of ^{14}C -based sedimentation rates dates the last three at ~41,000, ~52,000 and ~69,000 years (Fig. 1). Comparable AMS ages for the first three layers have been obtained from three other cores (Table 1).

One of the most unusual properties of the Heinrich deposits is the presence of prominent layers with 20% to 25% detrital limestone and dolomite, a composition markedly different from that in the ambient glacial sediment which consists mostly of quartz and feldspar, variable amounts of black volcanic glass and only a few per cent detrital carbonate (see for example Fig. 2). These distinctive layers have a well defined distribution in the northern Atlantic (Fig. 1; Table 3). In the southwest, all six Heinrich deposits contain a layer with abundant detrital carbon-

ate. To the north and east, however, H3 and H6 lack these layers, and their IRD resembles that of the ambient sediment. Still farther to the north and in the southernmost samples, none of the Heinrich deposits identified so far contain detrital carbonate. Hence, layers rich in detrital carbonate are confined to the belt of high IRD accumulation that formed during the last glaciation³ (Fig. 1).

Evidence of at least one source of the carbonate IRD has been found in the Labrador Sea (Fig. 1). Here, H1 and H2 contain thick layers rich in detrital carbonate (Tables 1-3) that was derived from nearby limestone and dolomite bedrock in Hudson Strait^{4,5} (Fig. 1). A source in eastern Canada is further supported by the marked westward increase in thicknesses of the carbonate-rich layers (Fig. 1). In addition, in the Dreizack cores the <2- μm carbonate-free fraction in H1, 2, 4 and 5 has an average K/Ar age of 897 Myr, contrasting with average ages of 442 Myr for the >2- μm ambient glacial sediment^{6,7} (Fig. 2). The older ages are evidence of detrital clay with large amounts of Precambrian material^{6,7}, ample sources of which occur in eastern Canada.

Two other features of the Heinrich deposits are worthy of mention. First, each layer rich in detrital carbonate accumulated rapidly. This is indicated by their sharp bases as revealed in X-radiographs of Dreizack^{8,9} and DSDP site 609 cores, and by increased fluxes of lithic grains comprising the carbonate IRD in H1 and H2 (Table 2). Foraminifera are well preserved in these layers and our flux estimates are not affected by dissolution. Second, all six Heinrich deposits contain evidence of surface water with extremely low temperature, low flux of planktonic foraminifera and low salinity. In each deposit, the planktonic species *Neogloboquadrina pachyderma* (left-coiled) dominates the foraminifera population (Fig. 2), indicating deep southward penetration of polar water. In DSDP site 609, the flux of planktonic foraminifera of size >150 μm dropped markedly during

TABLE 1 AMS radiocarbon dates for DSDP 609, HU75-55, V23-81 and V23-16

Core	Depth (cm)	Corrected age* (years)	Error (years)
DSDP 609 (G.b.)	63-65	11,980†	120
SDSP 609	64-66	11,180‡	190
DSDP 609	69-70	11,020‡	190
DSDP 609	73-75	12,350‡	220
DSDP 609	75-77	13,490‡	220
DSDP 609 (G.b.)	79-81	13,250‡	090
DSDP 609	84-85	14,590‡	230
DSDP 609	87-88	15,960‡	240
DSDP 609	90-91	16,360‡	150
DSDP 609	98-99	16,960‡	120
DSDP 609	105-107	18,940‡	220
DSDP 609	110-111	19,970‡	330
DSDP 609	111-112	20,550§	260
DSDP 609	112-113	21,110‡	220
DSDP 609	115-116	21,370‡	220
DSDP 609	118-120	22,380‡	340
DSDP 609 (G.i.)	139-141	25,260‡	440
DSDP 609	143-144	26,570‡	490
DSDP 609	147-149	26,170‡	310
DSDP 609	153-155	29,170†	660
DSDP 609	166-167	30,080‡	680
DSDP 609 (G.i.)	174-176	30,720†	730
HU75-55	81	13,235	190
HU75-55	116	14,610	105
HU75-55	181	19,455	210
HU75-55	250	21,105	240
V23-81	198-199	12,320§	220
V23-81	215-216	13,260§	210
V23-81	249-250	16,550‡	190
V23-81	264-265	18,530‡	240
V23-81	335-336	21,950‡	310
V23-81	340-341	23,090‡	370
V23-81	400-401	31,250‡	690
V23-81	418-419	30,150‡	770
V23-16	29-30	10,050‡	150
V23-16	54-55	15,370‡	220
V23-16	62-63	16,670‡	280
V23-16	144-145	23,400‡	390
V23-16	184-185	28,240‡	600
V23-16	207-208	31,230‡	880

* Corrected for assumed 400-yr difference between surface water carbon and atmospheric carbon. G.b. indicates analyses done on *G. bulloides*, G.i. indicates analyses done on *G. inflata*. All other measurements are on *N. pachyderma* (l.c.) H1 and so on are Heinrich deposits as defined by low foram concentrations (see for example Fig. 2). Hc refers to layers within Heinrich deposits that are rich in detrital carbonate IRD.

† Ref. 29.

‡ From analyses done in Zurich for this paper.

§ Ref. 19.

TABLE 2 Foraminifera and IRD sand flux estimates for DSDP 609

Depth interval (cm)	Radiocarbon time interval	Flux of foraminifera >150 μm (number $\text{cm}^{-2} \text{yr}^{-1}$)	Flux of IRD >150 μm (number $\text{cm}^{-2} \text{yr}^{-1}$)
64-74	1,270	29.31	3.78
74-76	1,140	5.39	2.58
76-84	1,100	7.40	10.50
84-87	1,370	6.34	3.14
87-90	400	27.57	11.49
90-105	2,580	26.87	6.07
105-110	1,030	9.89	4.78
110-112	1,140	4.45	2.85
112-115	260	8.53	20.15
115-118	1,010	9.78	5.25
118-139	2,880	17.24	4.17
139-143	1,310	5.58	1.05
143-153	2,600	1.97	2.13
153-166	910	17.79	9.10
166-174	540	22.76	7.02

Flux estimates were calculated using measured dry sample weights, grain and foram counts, radiocarbon age intervals and dry bulk density. Dry bulk density was calculated using wet bulk density and sediment porosity in ref. 20. Two equations were solved: $\text{flux} = \text{SI} [\sum (\text{pc}) \times \rho_{\text{dd}}] / \Delta t$, and $\rho_{\text{dd}} = \rho_{\text{wd}} - (\phi_s \times \rho_w)$, where SI is sampling interval, pc is particle count per gram. ρ_{dd} = dry bulk density, ρ_{wd} = wet bulk density, ϕ_s is sediment porosity, ρ_w is density of sea water, and Δt is time interval based on radiocarbon dates. Because layers with carbonate IRD have higher densities than ambient sediment and they lack foraminifera, the IRD flux values for those layers are a minimum. The decrease in foraminifera flux in the H layers is too large to be an artefact of variations in density and porosity of typical glacial sediment¹⁷.

FIG. 1 Location of cores containing Heinrich deposits (deposits with unusually large concentrations of IRD and low amounts of foraminifera) and IRD unusually rich in limestone and dolomite grains. Filled circles indicate that carbonate-rich IRD is present in all Heinrich deposits identified; half-filled circles indicate that carbonate-rich IRD is absent in some of the Heinrich deposits identified; open squares indicate that carbonate-rich IRD is absent in all of the Heinrich layers identified (see Table 3). The solid black pattern is the distribution of limestone and dolomite bedrock (including submerged rocks in Hudson Bay and Hudson Strait) and the thick solid line is the approximate maximum limit of ice sheets during the last glaciation. Cs are small exposures of carbonate rocks. The westward increase in thicknesses of layers rich in carbonate IRD, indicated by increase in the size of the circles, and the widespread occurrence of limestone and dolomite in eastern Canada and northwestern Greenland are evidence that the carbonate-bearing icebergs originated in the Labrador Sea and Scotian Shelf regions. Their transit completely across the Atlantic was made possible by extreme cooling of surface waters, as indicated by the high percentages of the polar planktonic foraminifera *N. pachyderma* in the Heinrich layers (Fig. 2). The arrows show possible paths of iceberg transport.

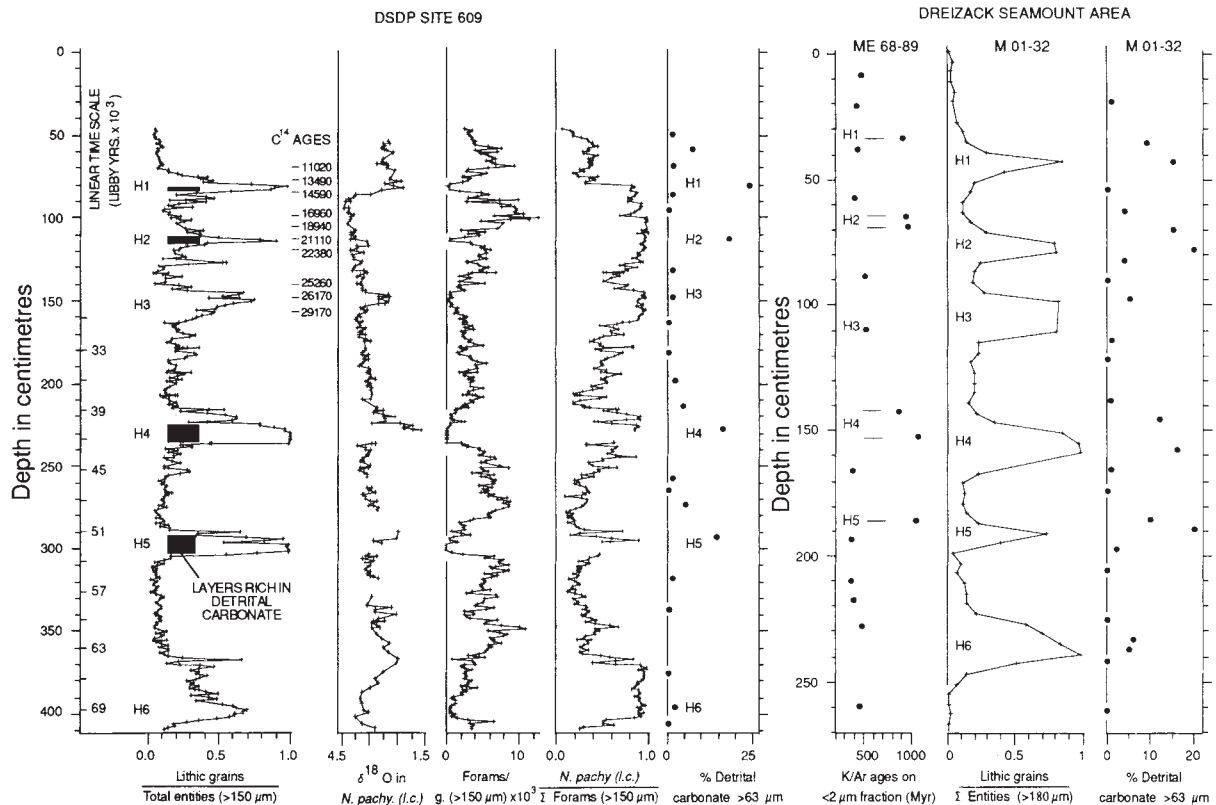
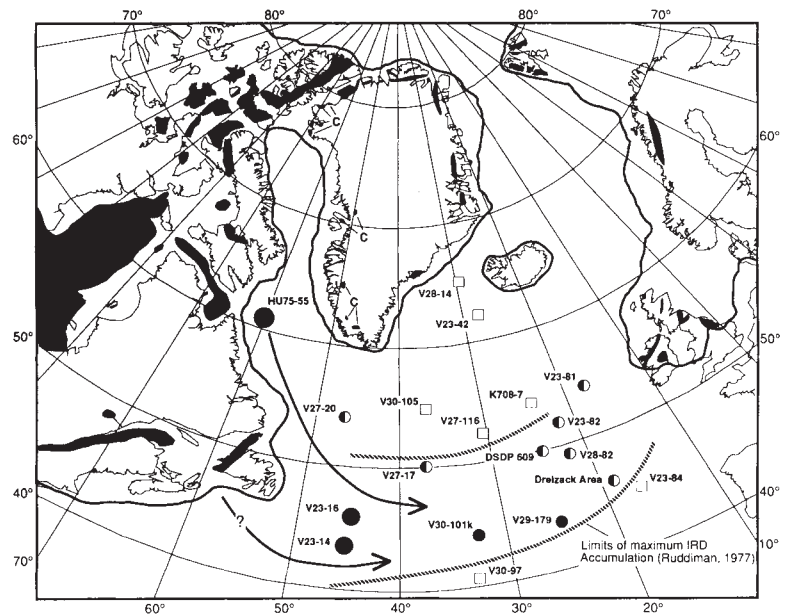


FIG. 2 Summary of radiometric, planktonic $\delta^{18}\text{O}$, foraminiferal and lithic measurements in cores from DSDP site 609 and from the Dreizack seamount area. The number of foraminifera size $>150\ \mu\text{m}$ per gram and the ratio of lithic grains to total entities $>150\ \mu\text{m}$ in cores from both localities defines the six deposits originally identified by Heinrich¹. Microscopic examination indicates that low foraminiferal concentrations also occur in material $<150\ \mu\text{m}$ in the Heinrich layers. AMS C^{14} ages are from Table 1. The linear timescale begins at 175 cm (30,720 C^{14} years, Table 1), and is based on a mean sedimentation rate of 6 cm per 1,000 years given by C^{14} dating and by orbital tuning²⁷. K/Ar ages from core Me 68–89 are from refs 6 and 7. The $\delta^{18}\text{O}$ measurements were made at Gif-sur-Yvette. ($\delta^{18}\text{O} = (\text{R}_{\text{sample}}/\text{R}_{\text{standard}} - 1)$, where standard is PDB.) The gaps in the ratio of *N. pachyderma* to total of foraminifera and in the $\delta^{18}\text{O}$ record in DSDP site 609 are intervals where the number of foraminifera is nearly zero. In these cores, four of the Heinrich deposits (H1, 2, 4 and 5) contain layers with unusually high percentages of carbonate IRD that was derived

from sources in eastern Canada (Fig. 1). The black bars in DSDP 609 are the positions of these layers in the core as defined by their distinctive appearance in X-radiographs. The detrital carbonate constitutes 20% to 25% of the IRD in both the Dreizack area and at DSDP site 609, determined by line counting the $>63\ \mu\text{m}$ fraction with a petrographic microscope. Based on visual estimates of DSDP site 609 material, detrital carbonate constitutes 40% to 60% of the $<63\ \mu\text{m}$ fraction. The detrital carbonate grains are rounded to subrounded and have fine-grained recrystallized textures. At one locality in the Dreizack area, core 244, oolites and fragments of echinoderms have been identified in the carbonate IRD. In addition, the layers rich in detrital carbonate have small amounts of fine-grained authigenic dolomite and unusually low porosities that average about one half that of typical deep sea muds (refs 8,9). The clay mineralogy of layers H1, 2, 4 and 5 in the Mount Dreizack cores also is unusual, consisting of abundant material from freshly derived acidic and metamorphic rocks and lacking the smectitic products of weathered basalts found in the ambient sediment (refs 6, 28).

TABLE 3 Thicknesses of detrital carbonate-rich layers within Heinrich deposits (H)

Core	Latitude (N)	Longitude (W)	Water depth (m)	H1		H2		H3		H4		H5		H6	
				d	t	d	t	d	t	d	t	d	t	d	t
HU75-55	61° 30'	58° 39'	2,500	90	40	210	69	*	*	*	*	*	*	*	*
V23-16	46° 00'	45° 03'	2,813	45	25	120	35	205	20	270	50	375	8	550	12
V23-14	43° 24'	45° 15'	3,177	*	*	55	10	120	20	150	15	265	25	380	15
V28-82	49° 27'	22° 16'	3,935	60	5	98	15	*	*	192	23	*	*	*	*
V30-101k	44° 06'	32° 30'	3,504	37	3	52	8	80	5	98	4	*	*	*	*
DSDP 609	49° 53'	24° 14'	3,884	82	1	113	5	150	0	229	10	296	10	400	0
V29-179	44° 01'	24° 32'	3,331	60	2	83	3	120	1	200	4	*	*	*	*
M01-32	47° 35'	28° 56'	4,070	42	3	78	7	105	0	157	12	192	5	236	0
V23-82	52° 35'	21° 56'	3,974	90	1	152	5	190	0	270	4	340	2	440	0
V23-81	54° 15'	16° 50'	2,393	220	1	329	3	383	0	495	3	584	2	760	0
V30-105	54° 31'	36° 30'	2,758	110	0	*	*	*	*	*	*	*	*	*	*
V27-17	50° 05'	37° 18'	4,054	33	0	49	5	70	2	95	3	*	*	147	2
V27-20	54° 00'	46° 12'	3,510	56	0	96	2	125	0	210	8	*	*	*	*
V27-116	52° 50'	30° 20'	3,202	43	0	65	0	*	*	*	*	*	*	*	*
V28-14	64° 47'	29° 34'	1,855	150	0	*	*	*	*	*	*	*	*	*	*
V23-42	62° 11'	27° 56'	1,514		NH	*	*	*	*	*	*	*	*	*	*
K708-7	53° 56'	24° 05'	3,502	40	0	100	0		NH		NH		NH	200	0
V30-97	41° 00'	32° 56'	3,371	70	0	*	*	*	*	*	*	*	*	*	*
V23-82	46° 00'	16° 55'	4,513	50	0	*	*	100	*	*	*	*	*	*	*

Core locations are given in Fig. 1. Thicknesses (t) and depths (d) are in centimetres; where layer rich in detrital carbonate is present, depths are at the approximate middle of that layer; otherwise depths are at the middle of the low foraminiferal zone. NH-No Heinrich deposit (that is, no deposit with low abundance of foraminifera) at expected depth. Data for identification and correlation of Heinrich deposits as follows: HU75-55, DSDP 609 (Table 1, refs 2, 4, 5); M01-32 (ref. 1); V23-81 (refs 19, 21 and W. F. Ruddiman, unpublished data from the National Geophysical Data Center (NOAA)); K708-7 (refs 21–23, and W. F. Ruddiman, NOAA unpublished data); V28-82 (% *N. pachyderma* from G.B. *et al.*, manuscript in preparation); V28-14 (refs 21, 23, 24); V23-16, V23-82, V23-83, V30-101k, V29-179, V30-105, V27-17, V27-20, V27-116, V23-42, V30-97 (refs 21, 23, 25, 26 and W. F. Ruddiman, NOAA unpublished data; and % *N. pachyderma* from G.B. *et al.*, manuscript in preparation). In V23-84 foraminiferal abundances have not been measured, but low coccolith zones occur at probable depths of H1 and H3 (ref. 26) and are taken as evidence of Heinrich deposits.

* No data available.

formation of at least the first three deposits (Table 2). This implies a reduction in the number of foraminifera living in the surface water, although we note that dissolution of foraminifera has occurred in H3. A reduction in the number of foraminifera in surface water could be associated with changes in primary productivity, but it could also reflect other changes in surface water, associated directly with temperature or salinity. Conspicuous decreases in planktonic $\delta^{18}\text{O}$ also occur in all six of the Heinrich deposits (Fig. 2). As the foraminiferal assemblages dictate cold conditions, sea surface salinities must have dropped markedly during deposition of each layer.

The evidence we have so far allows us to propose a tentative explanation for Heinrich layers and their unusual mineralogy. First, we note that the layers with abundant detrital carbonate lie within but do not extend through the Heinrich deposits (Fig. 2). This leads us to view Heinrich deposits as a composite of two processes. One was a drop in sea surface temperatures accompanied by reduction of the flux of planktonic foraminifera, accounting in part for the high lithic to foraminifera ratios originally noted by Heinrich¹. The other, indicated by rapid deposition of the carbonate-rich IRD, was a brief but massive discharge of icebergs, occurring within but not spanning the interval of low surface temperatures. The rapid deposition of this IRD overwhelmed the already reduced flux of foraminifera, producing layers within the Heinrich deposits nearly devoid of foraminifera (Fig. 2). We suggest that shortly after surface temperatures and foraminifera flux began to drop, ice streams in eastern Canada (and possibly in northwestern Greenland) began to advance rapidly, leading to massive calving as ice fronts reached maximum seaward positions. This led to the release of large amounts of glacier ice, much of it carbonate-bearing, into the Labrador Sea and perhaps into the Scotian shelf as well. Cold sea surface temperatures slowed melting and the icebergs drifted into the North Atlantic. Rapid transit of huge amounts of this ice eastward along the southern edge of the polar gyre led to rapid deposition of carbonate-rich IRD in a wide belt, reaching completely across the ocean during four of the events.

The absence of the carbonate-rich layers south of the IRD belt (Fig. 1) probably reflects complete melting of ice upon reaching warm water south of the glacial polar front³. Similarly,

the absence of carbonate-rich IRD in H3 and H6 in cores within the belt is probably a consequence of warmer sea surface temperature causing melting of the ice before it could traverse the width of the Atlantic. The absence of the carbonate-rich IRD in virtually all of the Heinrich deposits identified farther north is consistent with the glacial circulation pattern in which discharge of ice from the Labrador Sea was carried southwards and then in a broad arc northeastwards along the southern boundary of the polar gyre¹⁰.

Melting of the huge quantities of ice drifting across the North Atlantic must have been an important factor in the sudden lowering of surface salinities implied by the $\delta^{18}\text{O}$ measurements¹. Assuming a $\delta^{18}\text{O}$ of -35% for glacial ice¹¹ and ignoring the decrease in temperature implied by the foraminiferal speciation, the $\sim 1\%$ $\delta^{18}\text{O}$ changes require mixing only about 1 part iceberg meltwater with about 30 parts sea water. We note that the accompanying salinity drop is probably enough to shut down the North Atlantic's thermohaline circulation.

Perhaps the most important result of our study is its evidence of repeated, rapid advances of ice sheets in eastern North America and the questions it raises about the cause of those events. One explanation is stochastic surging², but that does not account for evidence of extreme cooling of surface water and reduced flux of planktonic foraminifera before and after the discharge of ice into the ocean. The same evidence makes it unlikely that the discharges of ice and melt products were the sole cause of the lowered sea surface temperatures. If atmospheric cooling accompanied the shift of polar water southward, the ice advances may have been forced by falling temperatures. The time intervals between the successive Heinrich layers, however, are shorter than orbital precession cycles (19 and 23 kyr), raising the question of what would have caused the atmospheric cooling. So far, only H1 and H2 have been linked directly to advances of ice streams, and only in the limited area of Hudson Strait^{4,5}. Our study raises questions of whether other parts of the Laurentide ice sheet discharged carbonate-bearing icebergs at the same times, whether there are glacial records of the older Heinrich events in eastern North America, and whether there were correlative advances of the Greenland, Fennoscandian and Barents Sea ice sheets^{12–18}. If so, those will be the first indications that large portions of Northern Hemisphere ice

sheets advanced synchronously on timescales less than those of the Milankovitch cycles, events that are not incorporated in current models of ice-sheet dynamics. □

Received 8 May; accepted 8 October 1992.

1. Heinrich, H. *Quat. Res.* **29**, 142–152 (1988).
2. Broecker, W. S., Bond, G., Klas, M., Clark, E. & McManus, J. *Clim. Dynam.* **6**, 265–273 (1992).
3. Ruddiman, W. *Geol. Soc. Am. Bull.* **88**, 1813–1827 (1977).
4. Andrews, J. T. & Tedesco, K. *Geology* (in the press).
5. Tedesco, K. & Andrews, J. T. *Eos* **72**(44), 271 (1992).
6. Jantschik, R. & Huon, S. *Ecol. Geol. Helv.* **85**, 195–212 (1992).
7. Huon, S., Jantschik, R., Kubler, B. & Fontignie, D. *Bull. Suisse Miner. Petrogr.* **21**, 275–280 (1991).
8. Jantschik, R. & Lohoff, R. thesis Univ. of Göttingen (1987).
9. Clausen, S., Jantschik, R. & Meischner, D. *Nachr. Dt. Geol. Ges.* **43**, 121 (1990).
10. Ruddiman, W. F. & McIntyre, A. *Science* **212**, 617–627 (1981).
11. Dansgaard, W. *et al. Science* **218**, 1273–1277 (1982).
12. Heinrich, R., Kasse, H., Vogelsang, E. & Thiede, J. *Mar. Geol.* **86**, 283–319 (1989).
13. Wolf, T. C. W. *Geomar. Rep.* **5**, 92 (1991).
14. Heinrich, R. in *Init. Rep. DSDP 104* (eds Eldholm, O., Thiede, J. & Taylor, E.) 189–232 (U.S. Govt. Printing Office, Washington, DC, 1989).
15. Vogelsang, E. *Ber. Sonderforschungsber.* **313**(32), 136 (1990).
16. Bischof, J. *Ber. Sonderforschungsber.* **313**(30), 127 (1991).
17. Kasse, H. *Ber. Sonderforschungsber.* **313**(24), 115 (1990).
18. Birgisdóttir, L. *Ber. Sonderforschungsber.* **313**(34), 112 (1991).
19. Broecker, W. S. *et al. Paleoceanography* **3**, 1–19 (1988).
20. Ruddiman, W. F. *et al. Init. Rep. DSDP 94*, 247–349 (U.S. Govt. Printing Office, Washington DC, 1987).
21. Ruddiman, W. F. & McIntyre, A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **35**, 145–214 (1981).
22. Ruddiman, W. F., Sancetta, C. D. & McIntyre, A. *Phil. Trans. R. Soc. Lond. B* **280**, 119–142 (1977).
23. Ruddiman, W. F., McIntyre, A., Niebler-Hunt, V. & Durazzi, J. T. *Quat. Res.* **13**, 33–64 (1980).
24. Kellogg, T. B. in *Climate Changes on a Yearly to Millennial Basis* (eds Möner, N. A. & Karlén, 123–133 (Reidel, Dordrecht, 1989).
25. Sancetta, C., Imbrie, K. J., Kipp, N. G., McIntyre, A. & Ruddiman, W. F. *Quat. Res.* **2**, 363–367 (1972).
26. McIntyre, A., Ruddiman, W. F. & Jantzen, R. *Deep-Sea Res.* **19**, 61–77 (1972).
27. Ruddiman, W. F., Raymo, M. E., Martinson, D. G., Clement, B. M. & Backman, J. *Paleoceanography* **4**, 353–412 (1989).
28. Grousset, F. & Chesselet, R. *Earth planet. Sci. Lett.* **78**, 271–287 (1986).
29. Broecker, W. S., Bond, G., Klas, M., Bonani, G. & Wolff, W. *Paleoceanography* **5**, 469–477 (1990).

ACKNOWLEDGEMENTS. We thank D. MacAyeal and G. Denton for comments on the manuscript, and J. Thiede and T. Wolf for information on possible Heinrich events in the Norwegian Sea. This research was supported in part by the NSF and NOAA. Work on the Dreizack cores was supported by a DFG grant. The planktic $\delta^{18}\text{O}$ have been measured with the support of the French CNRS, CEA and CEE(EPOCH). We thank the Ocean Drilling Program for permission to sample cores from DSDP site 609. Support for the core collection of Lamont-Doherty Geological Observatory is provided by the NSF and the office of Naval Research.

Primary production in the glacial North Atlantic and North Pacific oceans

Constance Sancetta

Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA

THE conditions controlling primary production are very different in the modern North Atlantic and North Pacific oceans¹, a difference that is reflected in the composition of diatom fossils in surface sediments. By contrast, I report here evidence that during the last glacial interval the diatom assemblage, and by extrapolation the primary production, was very similar in the two regions. The modern analogues of these assemblages occur in sediments of Baffin Bay and the Sea of Okhotsk, both highly productive seas where ice is present. I infer that during the last glacial interval plankton biomass was at least as high as it is today in the North Atlantic, and was as much as an order of magnitude higher in the North Pacific. The glacial assemblage occurs in lithologies dominated by ice-rafted detritus, which is generally believed to indicate the presence of icebergs^{2–6}. I hypothesize that the presence of numerous icebergs, possibly associated with sea ice, supported high production by physical mechanisms (such as turbulent mixing and enhanced density stratification) and/or biogeochemical ones (such as supply of major or trace nutrients).

In the modern world, primary production in the Pacific Ocean between 45° N and 60° N is dominated by production of picoplankton and is not limited by the availability of major nutrients. It has been suggested that picoplankton abundance is limited

by efficient microzooplankton grazing, and that diatom production may be limited by very low availability of a trace element such as iron^{1,7–9}. Coastal zone colour scanner (CZCS) imagery indicates a biomass of 0.3–0.5 mg C m⁻³ (ref. 10) and chlorophyll *a* concentrations rarely exceed 0.4 mg m⁻³ (ref. 9). The sediment diatom assemblage is dominated (60–80%) by *Denticulopsis seminae*¹¹, an endemic taxon which also dominates the diatom plankton assemblage¹².

In contrast, primary production in the same latitudes of the North Atlantic is characterized by a large spring bloom triggered by thermal stratification¹ and possibly seeded by transport from the coastal zone (R. Sambrotto, personal communication); summer production is nutrient-limited and may depend on recycled nutrients¹. Biomass estimated from CZCS is 1.0–10.0 mg C m⁻³ (ref. 10), at least an order of magnitude higher than that in the North Pacific, and chlorophyll *a* levels during blooms exceed 1 mg m⁻³ (ref. 1). Sediments underlying this area contain abundant spores of the genus *Chaetoceros* (40–60%) and a variety of sub-tropical taxa¹³. *Chaetoceros* are characteristic of the regional spring bloom (ref. 1 and A. Weeks and R. Williams, personal communication), whereas the sub-tropical taxa probably reflect transport by the West Wind Drift¹³. The difference in sediment diatom assemblages thus parallels the modern differences in seasonal conditions of primary production.

I have examined the diatom fossil assemblage in core records spanning the last 20,000 years using two cores from the North Pacific and one from the North Atlantic. Ocean Drilling Project Site 609 is located in the central North Atlantic Ocean (49°53' N, 24°14' W), and recently has been the focus of studies^{4,5} addressing the short intervals of massive ice-rafting (which created the so-called Heinrich deposits) during the last glacial period. I examined samples dating from about 9,000 to 23,000 yr BP, at intervals of 1,100 to 1,500 years (ages from accelerator mass spectrometry (AMS) ¹⁴C dating of planktonic foraminifera⁴). Samples from Heinrich zones were examined but not included in the quantitative analysis. Diatoms are extremely rare in these intervals, and it would be necessary to process large volumes of material to obtain sufficient numbers for statistically reliable results. Two cores from the North Pacific were examined at intervals of less than 1,000 years, covering the same period of time. Age control on these cores (PAR87a-02, 54°17' N, 149°36' W and RNDB-PC11, 51°4' N, 167°58' E) is based on correlation to nearby cores dated by AMS ¹⁴C analysis of planktonic foraminifera^{14,15}.

The taxonomic difference between the two oceans persists in samples younger than 15,000 yr BP, representing the Holocene and latest glacial interval. This suggests that, throughout that time, conditions of production in the two oceans were similar to those at present. In contrast, samples representing the full glacial interval are very similar in both oceans, with the diatom assemblage characterized by *Actinocyclus curvatulus* and *Thalassiosira latimarginata*. Samples from the Pacific cores are dominated by *T. latimarginata* (40–60%), with *A. curvatulus* as a secondary component (10–20%), whereas in the Atlantic core the two taxa are present in similar proportions (15–30%). *Chaetoceros* spores are also common in the Atlantic material (30–50%), implying persistences of the spring bloom.

Modern analogues of the assemblages can be found in the western marginal seas of the two oceans. Cluster analysis shows that the assemblage in the Pacific cores is most similar ($r > 0.78$) to modern core-top assemblages in the Sea of Okhotsk¹¹. The Atlantic assemblage of Site 609 seems to be similar to those reported for core-top material from Baffin Bay¹⁶. The latter are described as having abundant *Chaetoceros* spores (30–70%) and an average of 21% *A. curvatulus* and 13% *T. latimarginata*, values close to those observed in glacial samples from Site 609.

In modern Baffin Bay, biomass (estimated by CZCS) is about 1.0 mg C m⁻³ (ref. 10), and the average chlorophyll *a* concentration in spring is 1.7 mg m⁻³ (ref. 17). These values are within the range of those for the modern North Atlantic, from which

I conclude that plankton biomass during the glacial interval was as high as it is today, but was achieved under different physical/chemical conditions which favoured a different ecosystem. The modern Sea of Okhotsk has a range in biomass similar to that of the North Atlantic and Baffin Bay ($1.0\text{--}10.0\text{ mg C m}^{-3}$; ref. 10). This implies that plankton biomass in the glacial North Pacific was as much as an order of magnitude greater than it is today, being about the same as that in the North Atlantic at the same time and occurring under similar environmental conditions.

The glacial diatom assemblage in both oceans appears consistently in sediments dominated by unsorted lithic particles, ranging in size from clay to cobbles, and characterized by rock and mineral types derived from the continental interiors^{2,5}. These include metamorphic and igneous rocks such as gneiss, granite, limestone and amphibolite, and a wide suite of related minerals. Radiochemical dating on selected minerals yields ages in excess of 400 Myr (ref. 5); spores and palynomorphs typical of Palaeozoic plants have also been found (B. Cornet, personal communication). It is generally accepted that the occurrence of such material thousands of kilometres from land is the result of transport by icebergs²⁻⁶. Although sea ice may also transport lithic particles¹⁸⁻²⁰, their nature is very different as they consist either of volcanic aerosols or of fine-grained sedimentary particles derived from the continental shelves. The coincidence of the unusual diatom assemblage with ice-rafted detritus (IRD) suggests that the presence of abundant icebergs may have created or at least contributed to a distinctive environment in the surface waters of both oceans during glacial times.

Whether sea ice was also common in the glacial North Atlantic and North Pacific is an open question. In high-latitude regions of steep continental relief, such as Baffin Bay, sea ice and icebergs occur together close to land. Sea ice may also occur without common icebergs, as in the Sea of Okhotsk and the Bering Sea. Sea ice generally does not persist as far from land as do icebergs, however, so that in the regions under discussion it is possible that relatively little sea ice was present. The only known fossil indicators of sea ice are diatom taxa such as *Nitzschia cylindrica* and *N. grunowii*, which occur today in sea ice of both hemispheres²¹. These taxa are present in low numbers (less than 5% of the assemblage) in the glacial-age samples, suggesting that at least some sea ice was present.

The effects of sea ice on primary production have long been recognized²², but far less attention has been devoted to the effects of icebergs. The release of fresh water by the melting ice may lead to vertical mixing and changes in density stratification²³⁻²⁶. An upwelling effect can be identified hundreds of metres away from a drifting iceberg^{25,26}. Very large icebergs may have drafts over 200 m, which can penetrate the nutricline associated with the pycnocline and bring nutrient-rich waters to the surface²³. Abundant icebergs might, over a period of years, even give rise to a regional salinity minimum¹⁴. Support for this inference is found in the oxygen isotope record of planktonic foraminifera at Site 609, which indicates that marked declines in salinity occurred during Heinrich events⁵.

A second class of possible effects is biogeochemical changes resulting from nutrient input by the melting ice. Major nutrients such as nitrate, ammonia and silicate can be present in icebergs and in sea ice at concentrations an order of magnitude higher than in the ambient water^{27,28} and some evidence suggests that melting can enrich nutrient concentrations in waters near icebergs^{27,29}. Another possibility is the addition of trace elements, an idea originally proposed to explain the difference in productivity of coastal and oceanic waters³⁰ and recently expanded with the focus on iron⁷. If the diatom assemblage in the modern North Pacific is iron-limited, the glacial assemblage of large centric diatoms might reflect increased availability of iron, and perhaps a change in the ecosystem dynamics of the region^{8,9}. Dissolvable iron has been found in a sample of Antarctic glacial ice at concentrations an order of magnitude higher than in the

ambient water³¹. Concentrations in sea ice and in icebergs from the Northern Hemisphere are not known.

Thus, there are a variety of ways in which icebergs, possibly associated with sea ice, could have modified the environment of northern high-latitude surface waters during glacial periods. The significance of these effects depends greatly on the concentration and composition of the icebergs, both of which are unknown and which might vary by an order of magnitude or more. As a conservative estimate for the North Atlantic, if we assume a production of $1 \times 10^4\text{ km}^3$ of ice per year (J. Andrews, personal communication), with each iceberg having a volume of $1 \times 10^{-1}\text{ km}^3$, and take the total area affected by ice-rafting as $6.9 \times 10^6\text{ km}^2$ (ref 6), the spatial density would be only one iceberg per 1,000 km^2 —several orders of magnitude too small to have a significant effect. If, on the other hand, iceberg production was an order of magnitude greater during postulated surge intervals^{4,5} and if they were concentrated in a narrow zone along the palaeo-Polar Front^{3,6}, the spatial density would be much higher. No comparable estimate can be made for the North Pacific, where the production and flux of ice from the Cordilleran ice sheet has not been estimated. Judging by the occurrence of IRD in the Pacific³, the flux was probably lower than in the North Atlantic, and high densities occurred only within 1,000 km from shore. Sea ice may have been present off the coasts of the Aleutian and Kuril arcs.

Regardless of whether icebergs really do have an effect, the results of this study indicate that conditions in the North Atlantic and Pacific oceans were very similar during glacial periods, as also shown by radiolarian fossils^{32,33}, and that primary production/biomass was equal to (North Atlantic) and substantially greater than (North Pacific) it is today. This conclusion may not apply to the Heinrich deposits⁵, in which diatoms were too few to be quantified. If, however, the iceberg hypothesis is correct, it implies that high production resulting from high nutrient supply also occurred during these periods. All microfossils, including planktic foraminifera, are extremely rare in the Heinrich deposits⁵; it remains uncertain whether this rarity reflects a brief decrease in plankton production or dilution resulting from the massive influx of IRD.

The trends in diatom taxa have been found in other cores throughout the entire North Pacific above 45°N , repeated during each glacial interval for the last 700,000 years¹¹. Further work is required to verify the extent in space and time of these trends in the North Atlantic; it is possible that the conditions described existed only along the zone characterized by high IRD in which Site 609 is located^{5,6}. The conclusions concerning high biomass and inferred high primary production, if supported by other evidence, may help to account for the lower levels of atmospheric CO_2 observed in glacial-age ice-core samples³⁴. □

Received 5 June; accepted 28 October 1992.

1. Parsons, T. R. & Lalli, C. M. *Oceanogr. Mar. Biol. A Rev.* **26**, 317–359 (1988).
2. Bramlette, M. N. & Bradley, W. H. *U.S. Geol. Surv. Prof. Pap.* **196**, 1–34 (1942).
3. Kent, D., Opdyke, N. D. & Ewing, M. *Bull. geol. Soc. Am.* **82**, 2741–2754 (1971).
4. Broecker, W., Bond, G., Klas, M., Clark, E. & McManus, J. *Clim. Dynam.* **6**, 265–273 (1992).
5. Bond, G. *et al. Nature* **360**, 245–249 (1992).
6. Ruddiman, W. F. *Bull. geol. Soc. Am.* **88**, 1813–1827 (1977).
7. Martin, J. H., Gordon, R. M., Fitzwater, S. & Broenkow, W. W. *Deep-Sea Res.* **36**, 649–680 (1989).
8. Morel, F. M. M., Rueter, J. G. & Price, N. M. *Oceanography* **4**, 56–61 (1991).
9. Miller, C. B. *et al. Limnol. Oceanogr.* **36**, 1600–1615 (1991).
10. Feldman, G. *et al. Trans. Am. Geophys. Un.* **70**, 634–641 (1989).
11. Sancetta, C. & Silvestri, S. *Paleoceanography* **1**, 163–180 (1986).
12. Anderson, G. C., Lam, R. K., Booth, B. C. & Glass, J. M. *NOAA Spec. Rep.* **76**, 477–798 (1977).
13. Maynard, N. G. thesis, Univ. of Miami (1974).
14. Keigwin, L. D., Jones, G. A. & Froelich, P. N. *Earth planet Sci. Lett.* (in the press).
15. Zahn, R., Pedersen, T. F., Bornhold, B. D. & Mix, A. C. *Paleoceanography* **6**, 543–560 (1991).
16. Williams, K. M. *Mar. Micropaleont.* **10**, 327–341 (1986).
17. Subba Rao, D. V. & Platt, T. *Polar Biol.* **3**, 191–201 (1984).
18. Larssen, B. B., Elverhoi, A. & Aagaard, P. *Polar Res.* **5**, 313–315 (1987).
19. Wollenburg, I., Pirman, S. & Lange, M. *Eos* **69**, 1263 (1988).
20. Barnes, P. W., Kempe, E. M. & Reimnitz, E. *Eos* **69**, 1263 (1988).
21. Horner, R. in *Sea Ice Biota* (ed Horner, R.) 148–157 (CRC Press, Boca Raton, 1985).
22. Hart, T. J. *Discovery Rep.* **21**, 261–356 (1942).
23. Neshyba, S. *Nature* **267**, 507–508 (1977).
24. Jacobs, S. S., Huppert, H. E., Holdsworth, G. & Drewry, D. J. *J. geophys. Res.* **86**, 6547–6555 (1981).
25. Josberger, E. G. in *Iceberg Utilization* (ed. Hussey, A. A.) 245–264 (Pergamon, New York 1978).

26. Allison, I., Knowles, K. & Wright, S. *Iceberg Res.* **9**, 3–9 (1985).
27. Shulenberger, E. *Polar Biol.* **2**, 149–158 (1983).
28. Jacobs, S. S., Gordon, A. L. & Amos, A. F. *Nature* **277**, 469–471 (1979).
29. Appollonio S. *Science* **180**, 491–493 (1973).
30. Gran, H. H. *Int. Conseil Explor. Mer. Jour.* **8**, 343–358 (1932).
31. Martin, J. H., Gordon, R. M. & Fitzwalter, S. E. *Nature* **345**, 156–158 (1990).
32. Morley, J. J. *Quat. Res.* **20**, 374–386 (1983).
33. Morley, J. J., Hays, J. D. & Robertson, J. H. *Deep-Sea Res.* **29**, 1485–1499 (1982).
34. Barnola, J. M., Raynaud, D., Korotkevich, T. S. & Lorius, C. *Nature* **329**, 408–414 (1987).

ACKNOWLEDGEMENTS. I thank S. Jacobs, J. Andrews, P. Mayewski, C. Miller and J. Martin for encouragement and advice, as well as many Omnet subscribers for information on icebergs. Samples and information on age control were generously provided by L. Keigwin, B. Bornhold, T. Pedersen, M. Klas and G. Bond. L. Keigwin, J. Morley and W. Smith made helpful comments. North Pacific core top samples were provided by the Lamont-Doherty Geological Observatory Core Repository.

Changes in static stress on southern California faults after the 1992 Landers earthquake

Ruth A. Harris & Robert W. Simpson

US Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA

THE magnitude 7.5 Landers earthquake of 28 June 1992 was the largest earthquake to strike California in 40 years. The slip that occurs in such an earthquake would be expected to induce large changes in the static stress on neighbouring faults; these changes in stress should in turn affect the likelihood of future earthquakes. Stress changes that load faults towards failure have been cited as the cause of small^{1–5}, moderate⁶ and large⁷ earthquakes; conversely, those that relax neighbouring faults have been related to a decrease in seismicity⁵. Here we use an elastic half-space model⁸ to estimate the stress changes produced by the Landers earthquake on selected southern California faults, including the San Andreas. We find that the estimated stress changes are consistent with the triggering of four out of the five aftershocks with magnitude greater than 4.5, and that the largest changes (1–10 bar), occurring on part of the San Bernardino segment of the San Andreas fault, may have decreased the time to the next magnitude 8 earthquake by about 14 years.

We evaluated the effects of the Landers earthquake on selected southern California faults (Fig. 1) by simulating the earthquake rupture with rectangular dislocation surfaces in an elastic half-

space⁸. The ten rectangles used for the model rupture were assigned right-lateral strike-slip offsets ranging from 1 to 5 m based on observations of surface slip. The calculated stress changes were resolved onto other rectangular surfaces, representing the nearby faults of interest, with centres at mid-seismogenic depths of 7.5 km (less for dipping rectangles). The three components of resolved stress change are the horizontal shear component, $\Delta\sigma_1$, the vertical or dip-parallel shear component, $\Delta\sigma_2$, and the normal component, $\Delta\sigma_3$ (positive for extension).

We also calculated the change in the Coulomb failure function, ΔCFF , to assess how the Landers-induced static-stress changes have increased or decreased the likelihood of subsequent earthquakes. ΔCFF is defined⁵ as

$$\Delta CFF = \mu(\Delta\sigma_3 + \Delta P) + \Delta\sigma_s = \mu'\Delta\sigma_3 + \Delta\sigma_s$$

where μ and μ' are the coefficient and apparent coefficient of friction, respectively. ΔP is the change in the pore-fluid pressure and $\Delta\sigma_s$ is the change in shear stress in the fault-slip direction. It is possible that immediately after the Landers event the change in pore-fluid pressure counteracted $\Delta\sigma_3$ to a degree⁹ so that the apparent coefficient of friction was low (~ 0.2)⁵. If the pore fluids then drain and the pore pressure re-equilibrates with time, the apparent coefficient of friction should rise to higher values (~ 0.8) more consistent with laboratory observations. For faults that are expected to slip in strike-slip earthquakes, such as the Garlock and San Jacinto faults, $\Delta\sigma_s = \Delta\sigma_1$ and the two stress components that determine ΔCFF are $\Delta\sigma_1$ and $\Delta\sigma_3$. For faults that slip in oblique-slip earthquakes, $\Delta\sigma_s$ includes both $\Delta\sigma_1$ and $\Delta\sigma_2$, so that $\Delta\sigma_1$, $\Delta\sigma_2$ and $\Delta\sigma_3$ all contribute to the change in Coulomb failure.

In our models, the largest static-stress changes on nearby faults occurred on the Calico and Lenwood faults (Fig. 1), which lie respectively to the east and west of Landers. For these two faults, left-lateral $\Delta\sigma_1$ of up to 17 bar indicates a relaxation relative to their expected right-lateral motion. The next-largest stress changes occurred along the San Bernardino segment of the San Andreas fault. This dipping segment, which is expected eventually to break in a magnitude 7.5–8 earthquake¹⁰ received horizontal and dip-parallel shear-stress changes of up to 6 bar (Fig. 2). The eastern half of the San Bernardino segment was also extended, and for Coulomb-type failure criteria (Fig. 3) this unclamping would be expected to bring the segment closer to failure for either right-lateral or right-lateral reverse-slip earthquakes. Although the picture is complex and varies with

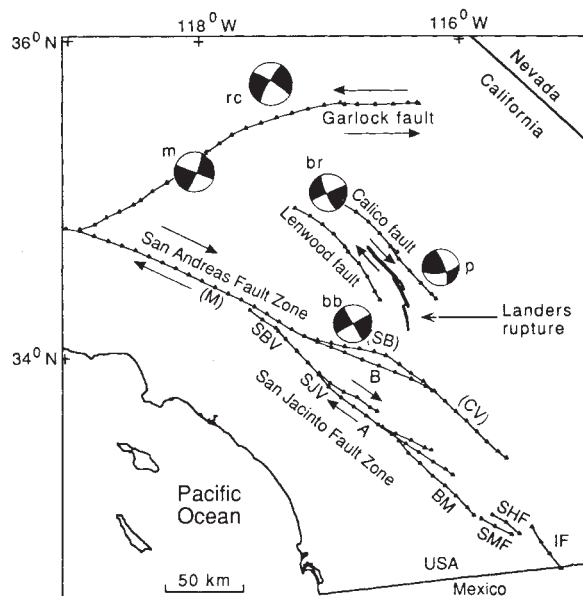


FIG. 1 Geological setting of the magnitude 7.5 Landers earthquake of 28 June 1992, which ruptured 70 km of the Johnson Valley, Homestead Valley, Emerson and Camp Rock faults in the eastern Mojave shear zone of southern California. This earthquake produced up to 5 m of right-lateral strike-slip offset on the trace of the Emerson fault (J. J. Lienkaemper and K. R. Lajoie, personal communication) with an average offset of approximately 3 m over its rupture length. Surface rupture is represented by the dark line-segments. The Big Bear (bb), Ridgecrest (rc), Pisgah (p), Mojave (m) and Barstow (br) epicentres (L. M. Jones, personal communication) are denoted by their respective focal mechanisms. Of the latter five earthquakes, four may have been triggered by the Landers earthquake static stress changes (Fig. 3). The San Andreas (SAFZ) and San Jacinto (SJFZ) fault zones are divided into segments¹⁰. From northwest to southeast the SAFZ is divided into the Mojave segment (M), the San Bernardino segment (SB) and the Coachella Valley segment (CV). From northwest to southeast the SJFZ is divided into the San Bernardino Valley segment (SBV), the San Jacinto Valley segment (SJV), the Anza segment (A) and the Borrego Mountain segment (BM). SMF and SHF are the Superstition Mountain and Superstition Hills faults, and IF is the Imperial fault. Small triangles show where the faults were divided into rectangular patches, approximately 10 km long by 15 km deep, onto which the stress changes were resolved (oblique view in Fig. 3). Patches representing the SB region of the SAFZ and the Banning fault¹⁸ (B) dip 50–70 degrees to the northeast^{19,20}. This represents the fault dips at approximately 6 km depth on the San Bernardino segment of the SAFZ and on the Banning fault.

location, it is clear that parts of the San Bernardino segment of the San Andreas fault are at more dangerous stress levels (by 1–10 bar) than they were before the Landers earthquake, and that time-delayed effects such as the re-equilibration of pore fluids⁹ may increase the proximity to failure.

Four out of five of the moderate-to-large-magnitude aftershocks of the Landers earthquake are consistent with triggering by static-stress changes induced by the Landers earthquake, although we cannot rule out the possibility that these aftershocks were triggered primarily by the dynamic stress changes that accompanied the passage of the seismic waves from the Landers shock. The magnitude 6.6 Big Bear earthquake of 28 June 1992,

the M_L 4.7 Ridgecrest earthquake of 30 June, the M_L 4.6 Barstow earthquake of 5 July, and the M_L 5.6 Mojave earthquake of 11 July (L. Jones, personal communication) all nucleated in regions, and on fault planes, brought closer to failure (Fig. 3) by the Landers earthquake (M_L is local magnitude). The M_L 5.5 Pisgah earthquake of 5 July (Fig. 1) is the only aftershock that does not seem to match the predicted static-stress changes. This exception suggests that there must be additional variables or other explanations than changes in static stress for the triggering of at least some aftershocks. Triggering by dynamic stresses has been mentioned already. Another possibility might be that the fault is weakened by the passage of the seismic waves, so that it can fail sooner, even though the static stresses acting on it might have been relaxed (J. H. Dieterich, personal communication).

The Landers earthquake induced much larger static-stress changes on the San Bernardino segment of the San Andreas fault than those experienced by the five aftershocks. This raises the question of why the San Andreas fault did not rupture immediately after the Landers event. Explanations include the possibility that the San Bernardino segment is very strong so that rupture does not nucleate in that segment. Although the San Bernardino region may participate in magnitude 7.5–8 earthquakes, these earthquakes may nucleate in other segments and propagate dynamically into the San Bernardino region. Another possibility is that even with the addition of the static-stress changes induced by the Landers earthquake, the San Bernardino region is many years from failure. Failure may be brought closer, however, as the pore fluids drain and the large normal-stress changes (Fig. 2) become more effective.

One technique for assessing the relative effects of the static-stress changes generated by the Landers earthquake is to compare the stress changes to the yearly stress load, which we estimated using a dislocation model that forced vertical dislocations extending from 15 to 100 km depth under the major faults to slip at the tabulated long-term slip rates. This model yields estimates of long-term loading rates (bars yr^{-1}) on the upper seismogenic reaches of these faults that can then be compared with the Landers static stress changes. The resulting years of advance and delay are necessarily approximate because of the simplicity of the model, especially for the important dipping

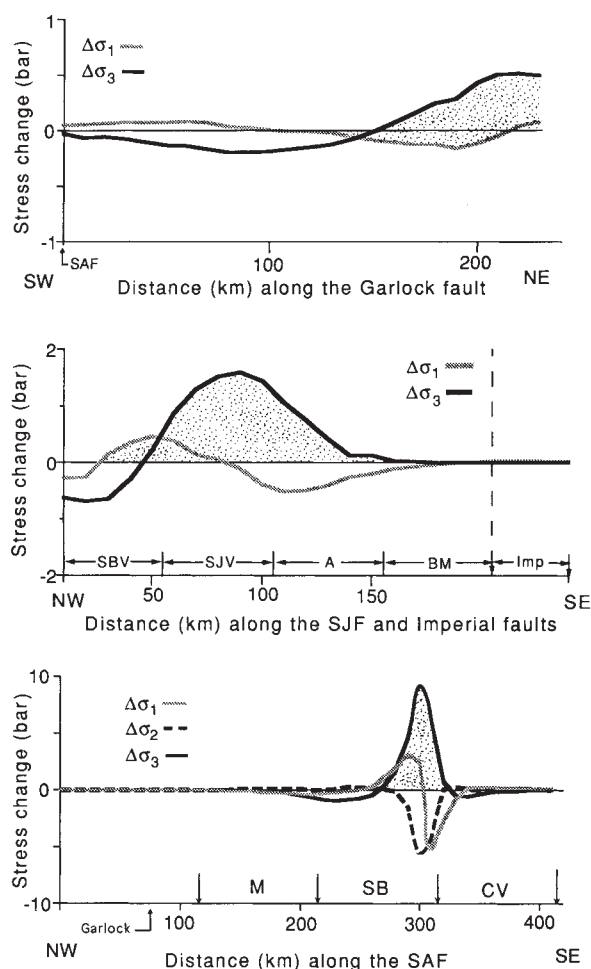


FIG. 2 Static-stress changes generated by the Landers earthquake, resolved onto four California faults and plotted as a function of distance along the strike of these faults. $\Delta\sigma_1 > 0$ for right-lateral shear, $\Delta\sigma_2 > 0$ for up-dip shear, and $\Delta\sigma_3 > 0$ for tension. Note that the scales vary for the three plots. Shading shows changes that bring the four faults closer to strike-slip failure. These plots illustrate some of the stress changes that contribute to Fig. 3. *a*, Static-stress changes along the left-lateral strike-slip Garlock fault. The eastern half of the fault was stressed towards failure by the Landers earthquake, with tension of up to 0.5 bar. The western half was moved away from failure by both a clamping of the fault and added right-lateral horizontal shear. *b*, Static-stress changes along the right-lateral strike-slip San Jacinto and Imperial faults. The San Jacinto fault experienced tension of up to 1.7 bar in the San Jacinto Valley region. The consequences of this change in the normal stress may become more evident over time. This effect could be monitored by observing changes in microseismicity rates along the SJV segment. *c*, Static-stress changes along the right-lateral strike-slip San Bernardino fault. The San Bernardino segment of the San Andreas fault may slip in right-lateral reverse-slip earthquakes, if the magnitude 5.9 1986 North Palm Springs earthquake²¹ which ruptured the parallel Banning fault is indicative of earthquakes in this dipping fault region. This implies that $\Delta\sigma_2$ plays an important role in determining failure conditions for the San Bernardino segment of the San Andreas.

TABLE 1 Effects of Landers earthquake on large Californian faults, relative to the long-term slip rate

Fault-zone segment ¹⁰	Long-term slip rate ^{11–15} (mm yr^{-1})	Date of last earthquake ¹⁰	Expected [†] recurrence time ¹⁰ (yr)	Years of advance*
San Andreas fault zone	30			
Mojave (M)		1857	162	–0.3
San Bernardino (SB)		1812(?)	198	14
Coachella Valley (CV)		1680 ± 20	256	3
San Jacinto fault zone	10			
San Bernardino Valley (SBV)		1890 (?)	203	6
San Jacinto Valley (SJV)		1918	184	5
Anza (A)		1774 ± 26‡	220–320‡	–2*
Borrego Mountain (BM)		1968	189	<0
Imperial fault	20	1979	44	0.2

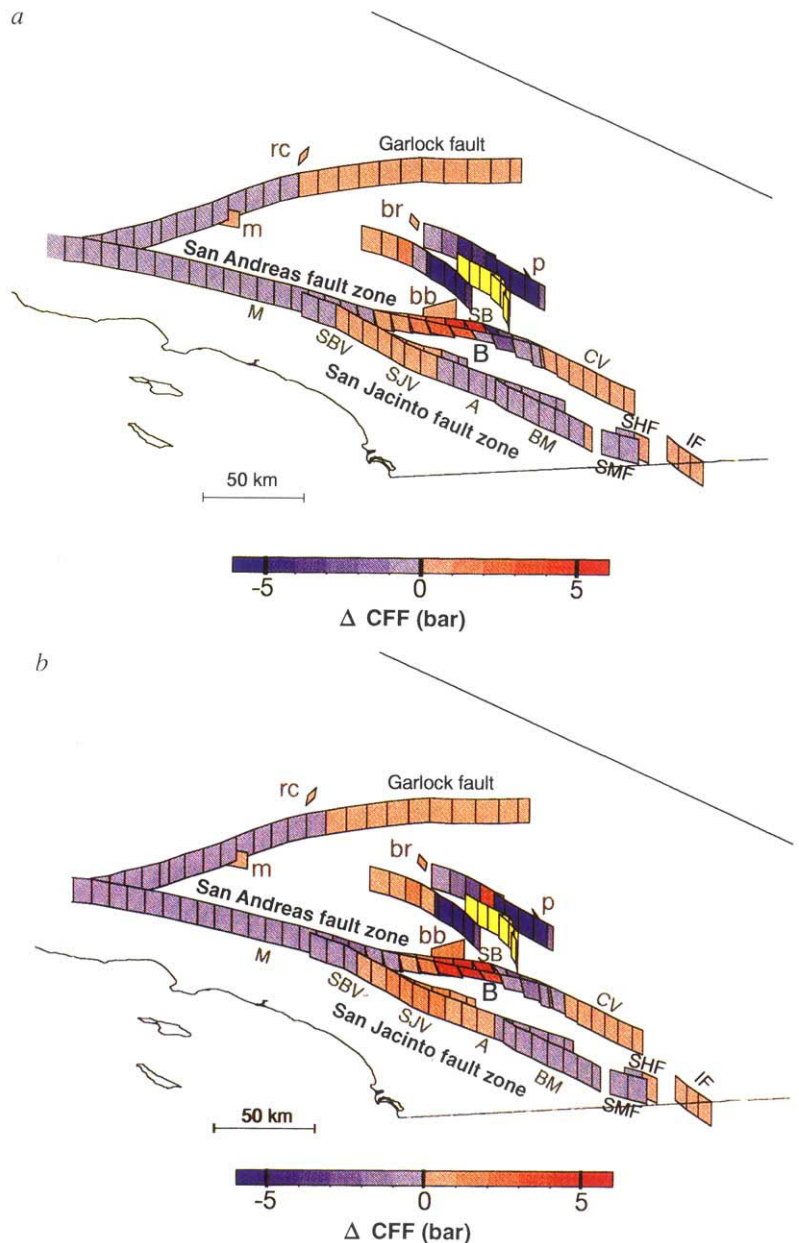
The Calico and Lenwood faults have not been included in this tabulation, because little is known at present about their long-term slip rates.

* Calculated by dividing $\Delta\sigma_1$ owing to the Landers event by the horizontal shear stress owing to the long-term fault slip rates. For the Anza segment, the advance also depends on the coefficient of friction used in the Coulomb failure function because the segment tends towards failure only if $\mu'\Delta\sigma_3$ exceeds the left-lateral change in the shear stress. The Anza segment was delayed for $\mu' = 0.2$, but $\mu' = 0.8$ advances the fault segment and increases the likelihood of failure (Fig. 3).

† Based on the time-predictable model¹⁶.

‡ From ref. 17.

FIG. 3 Oblique view from the southwest, showing Δ CFF produced by the Landers rupture (yellow rectangles) on selected southern California faults and on fault planes of five magnitude >4.5 aftershocks. The selected faults were chosen either because of their potential hazard¹⁰ or because of their proximity (Calico and Lenwood) to the Landers rupture. For the selected faults, Δ CFF was calculated by assuming that these faults will fail in right-lateral strike-slip earthquakes, except for the Garlock fault which was assumed to fail in left-lateral earthquakes. For the five aftershocks, Δ CFF was calculated by picking a fault planet and using the slip direction inferred from the focal mechanism. The five aftershocks probably occurred on pre-existing fault planes, although none occurred on our selected faults. Static stress changes on four of the five aftershock fault planes are in a direction to have encouraged these earthquakes. Red colours (Δ CFF > 0) imply that the fault was loaded closer to failure by the static stress changes from the Landers earthquake, blue colours (Δ CFF < 0) imply that the fault was brought farther from failure. *a*, Δ CFF values for $\mu' = 0.2$, which may best represent the change in failure immediately after the earthquake⁵. *b*, Δ CFF values for $\mu' = 0.8$, which may best represent either the dry case²² or the situation after enough time has passed for the pore pressures to re-equilibrate⁹. Parts of the San Bernardino (SB) segment are also loaded toward North Palm Springs-style right-lateral/reverse oblique slip failure, which is not shown in this figure, but which may be inferred from Fig. 2c.



part of the San Andreas fault in the San Bernardino segment, where the loading must be far more complex than we have represented in our model. Nonetheless, the results compare favourably with estimates made using other assumptions and offer at least a guide to expected changes. We find that the Coachella Valley segment of the San Andreas fault zone was brought closer to failure in some places by up to three years, and that the Imperial fault was also advanced towards failure. The San Bernardino Valley and San Jacinto Valley segments of the San Jacinto fault zone were advanced by six and five years respectively. The fault segment that was most advanced by the Landers earthquake is the San Bernardino segment of the San Andreas fault zone. Parts of it show a 14-year advance for right-lateral strike-slip earthquakes, and calculations also demonstrate an advance for right-lateral/reverse oblique-slip earthquakes on parts of this dipping fault segment. As shown in Table 1, however, there are also segments of the San Andreas and San Jacinto fault zones on which future earthquakes have been delayed by the Landers earthquake.

Although static stress changes have been used in a number of studies to explain patterns of aftershocks and the triggering of new shocks, the role of static stress changes in advancing or delaying future moderate to large earthquakes is still not clear.

The apparent correlation of microseismicity rate changes on San Francisco Bay area faults with calculated static stress changes after the 1989 Loma Prieta earthquake⁵ suggests that nearby faults are indeed feeling the stresses that we calculate. We plan to search for a similar post-Landers correlation in southern California as the data become available. Perhaps the most convincing evidence that static stress changes do affect the timing of future large earthquakes is negative: the 1906 San Francisco earthquake is calculated to have relaxed static stresses on most of the important faults in the Bay region. For 73 years after 1906, there was only one earthquake (in 1911) on Bay area faults with magnitude greater than 6 (ref. 23). This seismic quiescence ended in 1979 with a series of moderate earthquakes on the Calaveras fault. The location and timing of these earthquakes agree well with the calculated erosion of the 1906 relaxation effect by long-term tectonic loading. In southern California, the predicted delays on the Mojave and Anza segments, if fulfilled, will offer some additional support for the role of static stress changes, although these delays are too small to offer definitive evidence. In the long run, the best confirmation that static stress changes can affect the timing of future large earthquakes will probably come from studies of the interactions of large earthquakes over many earthquake cycles. □

Received 5 October; accepted 27 October 1992.

1. Das, S. & Scholz, C. H. *Bull. seismol. Soc. Am.* **71**, 1669–1675 (1981).
2. Stein, R. S. & Lisowski, M. *J. geophys. Res.* **88**, 6477–6490 (1983).
3. Oppenheimer, D. H., Reasenber, P. A. & Simpson, R. W. *J. geophys. Res.* **93**, 9007–9026 (1988).
4. Das, S. *Nature* **357**, 150–153 (1992).
5. Reasenber, P. A. & Simpson, R. W. *Science* **255**, 1687–1690 (1992).
6. Erickson, L. L. thesis, Stanford Univ. (1986).
7. Hudnut, K. W., Seeber, L. & Pacheco, J. *Geophys. Res. Lett.* **16**, 199–202 (1989).
8. Okada, Y. *Bull. seismol. Soc. Am.* **82**, 1018–1040 (1992).
9. Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 1990).
10. Working Group on California Earthquake Probabilities *U.S. Geol. Surv. Open-File Rep.* 88–398 (1988).
11. Keller, E. A., Borkowski, M. S., Korsch, R. J. & Shlemon, R. J. *Geol. Soc. Am. Bull.* **93**, 46–56 (1982).
12. Weldon, R. & Sieh, K. E. *Geol. Soc. Am. Bull.* **96**, 793–812 (1985).
13. Sieh, K. E. & Williams, P. L. *J. geophys. Res.* **95**, 6629–6645 (1990).
14. Sharp, R. V. *J. geophys. Res.* **86**, 1754–1762 (1981).
15. Clark, M. M. *et al.*, *U.S. Geol. Surv. Open-File Rep.* 84–106 (1984).
16. Shimizaki, K. & Nakata, T. *Geophys. Res. Lett.* **7**, 279–282 (1980).
17. Klirger, R. E. thesis, San Diego State Univ. (1989).
18. Matti, J. C., Morton, D. M. & Cox, B. F. *U.S. Geol. Surv. Open-File Rep.* 85–365 (1985).
19. Nicholson, C. & Lees, J. M. *Geophys. Res. Lett.* **19**, 1–4 (1992).
20. Nicholson, C., Seeber, L., Williams, P. L. & Sykes, L. R. *J. geophys. Res.* **91**, 4891–4908 (1986).
21. Jones, L. M., Hutton, K., Given, D. A. & Allen, C. R. *Bull. seismol. Soc. Am.* **76**, 1830–1837 (1986).
22. Byerlee, J. D. *Pure appl. Geophys.* **116**, 615–626 (1978).
23. Sykes, L. R. & Jaumé, S. C. *Nature* **348**, 595–599 (1990).

ACKNOWLEDGEMENTS. We thank L. Jones for the focal mechanisms of the post-Landers seismicity, R. Stein and G. King for discussion, C. Nicholson for insight into the San Bernardino region of the San Andreas fault, and P. Reasenber, M. Lisowski and D. Jackson for critically reading the manuscript. R.A.H. was supported by a National Research Council associateship at the USGS. We thank L. R. Sykes and S. C. Jaumé as well as R. S. Stein, G. C. P. King, and J. Lin for preprints of their stress studies of the Landers earthquake.

Sibling rivalry and brood sex ratios in polyembryonic wasps

Miodrag Grbić, Paul J. Ode & Michael R. Strand*

Department of Entomology, 237 Russell Laboratories, 1630 Linden Drive, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

FEMALE-BIASED sex ratios are predicted under local mate competition where offspring of one or a few females mate in isolated subpopulations¹. Haplodiploid sex determination allows parasitic wasps to bias their sex ratios by controlling fertilization², but parental control is not solely responsible for the broods of polyembryonic wasps. The eggs of these parasites divide to form many offspring and often two larval morphs^{3,4}. Precocious larvae die without pupating while reproductive larvae become adults. *Copidosoma floridanum* (Hymenoptera: Encyrtidae) produces 1,000–1,400 offspring per host and most broods contain both sexes (mixed)⁵. Mixed broods are female-biased and wasps mate before dispersing, but these broods develop from the mother laying one male and one female egg and adult males probably obtain additional matings away from the host^{5,6}. Thus, intersexual conflict may exist over the number of offspring to produce per host. Here we show that precocious larvae in mixed broods are predominantly female and that they bias the sex ratio by killing males.

C. floridanum is a small wasp that parasitizes the eggs of the moth *Trichoplusia ni*. Wasps produce all-male or all-female broods by laying one egg per host and mixed broods by laying two eggs (one male and one female)⁵. After parasitism the moth host egg hatches and the larva develops to its final (fifth) instar. During this period, the *C. floridanum* egg(s) proliferates clonally, forming a mass of embryos called a polygerm^{7–9}. Some embryos (<50) initiate morphogenesis during the host first-fourth instar and form precocious larvae, which have enlarged mandibles and die at the end of host development. The remaining embryos (>1,000) develop in the host fifth instar and form reproductive larvae which consume the host, pupate and synchronously emerge as adult wasps.

Examination of 58 *C. floridanum* broods from two cabbage fields in Wisconsin indicated that 18% were all-male, 15% were all-female and 67% were mixed. The mean $\pm$ s.e. number of

TABLE 1 Attack of brother and sister polygerms by precocious larvae

<i>In vivo</i>			
Sex of labelled polygerm injected into the host	<i>N</i>	Mean percentage of precocious larvae labelled per host	
Male	28	34	<i>P</i> < 0.0001*
Female	29	7	
<i>In vitro</i>			
	Attacked (%)	No response (%)	
Male versus female polygerm (<i>N</i> = 33)			
Male	63	37	<i>P</i> = 0.006†
Female	18	82	
Fat body versus testes (control) (<i>N</i> = 20)			
Fat body	5	95	<i>P</i> = 0.756†
Testes	5	95	

Precocious larvae attacked male polygerms significantly more often than female polygerms *in vivo* and *in vitro*. All experiments were conducted using hosts containing *C. floridanum* that were full sibs. All-male and female broods arise from a female laying one egg and are thus clonal. By mating females from one brood with males from another, all resulting progeny are brothers and sisters. *In vivo* interactions in the host cannot be observed directly so we ran tracer experiments. Male or female polygerms labelled with CFSE¹⁰ (see Fig. 2 legend) were injected into hosts containing a female brood using a glass needle mounted on a micromanipulator. To control for the possibility that some precocious larvae were labelled nonspecifically, wing disks from host fourth instar larvae (which are similar in size to a polygerm) were labelled and injected into hosts ($N = 20$) as described previously. No precocious larvae were labelled in these experiments. As a second control, no precocious larvae contained tracer when polygerms ($N = 20$) were labelled and injected into unparasitized hosts, indicating that larvae labelled in a host are not from the polygerm that was injected. Choice tests were conducted *in vitro* in 3- μ l culture wells filled with SF-900 medium (GIBCO). Precocious larvae, polygerms and host tissues (control) were collected from fourth instar hosts immediately before the experiment. A female precocious larva was placed in the culture well with a brother and sister polygerm and the sex of polygerm that was attacked was recorded. Assays lasted for 30 min and an attack was defined as the larva gripping a polygerm with its mandibles for more than 1 min. When this occurred, consumption of tissue by the larva was readily visible. Specificity of attacks on polygerm was investigated by conducting choice tests using host fat body and testes.

* $F_{1,56} = 29.9$ using arcsin transformed data.

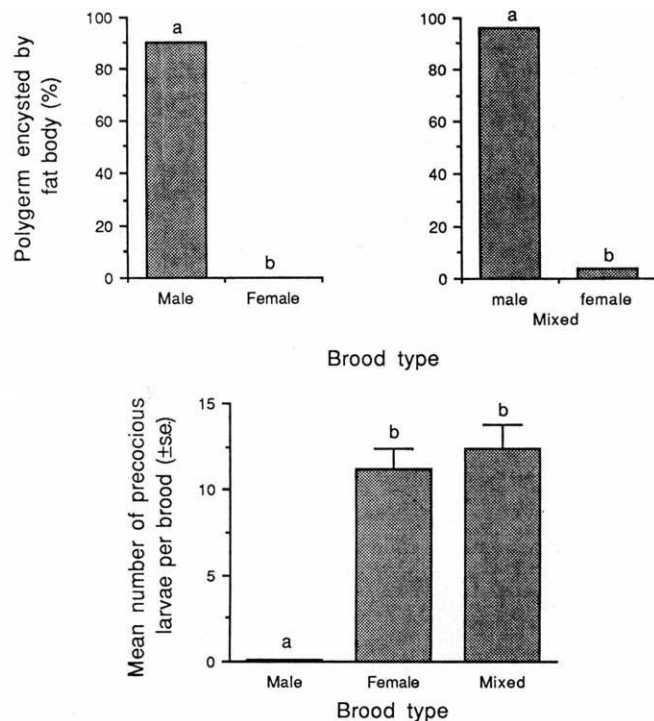
† Fisher's exact test.

wasps per host was $1,385.2 \pm 79.2$, with no differences in clutch size between the three brood types ($F_{2,54} = 2.5$, $P > 0.09$). The sex ratio of mixed broods was strongly female-biased (0.12 ± 0.03). Laboratory studies indicated that these broods were almost certainly of clonal origin because wasps from the field produced single-sex and mixed broods by laying one or two eggs and avoided laying in parasitized hosts. Mixed broods were also female-biased (0.09 ± 0.02 ; $N = 50$), indicating that the shift in sex ratio occurs during offspring development.

The sex ratio of mixed broods could arise for at least three reasons: (1) differential mortality due to exploitation competition with female larvae developing more rapidly than males; (2) differential mortality due to interference competition between larvae and/or embryos; and (3) differential proliferation of embryos. The first possibility is eliminated because development time of males and females is the same^{5,7}. Interference competition between reproductive-morph larvae is also eliminated because these larvae do not fight⁵. As the number of male and female wasps produced in single-sex broods does not differ, possibility (3) is also unlikely, although proliferation of male embryos could change in the presence of females.

To determine whether the sex ratios of mixed broods result from differential mortality of male embryos, we compared development of males and females. The most notable difference was that in all-female broods the polygerm was suspended freely in the host thorax, but in all-male broods the polygerm was usually encysted by fat body tissue and located in the abdomen

* To whom correspondence should be addressed.



(Fig. 1). We also observed that through the host fourth instar, 8–16 precocious larvae were present in female broods but precocious larvae were usually absent in male broods (Fig. 1). Similar differences were observed in mixed broods, suggesting sexual asymmetries existed in *C. floridanum* development (Fig. 1). Injection of brother or sister polygerms labelled with a vital tracer¹⁰ into hosts containing a sister brood indicated that precocious larvae preferentially killed males (Table 1; Fig. 2). *In vitro* assays confirmed these observations and suggested that attacks on polygerm were directed because precocious larvae rarely attacked host tissues in choice assays (Table 1).

Local mate competition favours a female-biased sex ratio because a reduction in the number of sons means less competition for mates and an increase in the number of daughters means more potential mates for sons^{11–14}. As males in mixed broods mate with sisters on the host before dispersing^{5,6}, conditions associated with local mate competition exist for *C. floridanum*. Males in a mixed brood should be selected to maximize the total number of females inseminated and no genetic conflict of interest will exist, provided males have the opportunity to mate only their sisters. But if males can obtain matings away from

FIG. 1 The percentage of polygerms encysted by fat body (top) and the mean number of precocious larvae per host (bottom) for all-male, all-female and mixed broods. Twenty hosts are examined for each type of brood. Bars labelled with different letters a and b are significantly different from one another at $P \leq 0.05$. Data were analysed by contingency table analysis (top) or an analysis of variance followed by Tukey–Kramer HSD multiple comparison procedure (bottom)²⁴.

METHODS. Two-hour-old host eggs were parasitized individually and the host larva was allowed to develop to the fourth instar on artificial medium at $27^\circ \pm 1^\circ$ with a 16-h light, 8-h dark photoperiod. Hosts were then dissected in physiological saline and the polygerm(s) and precocious larvae located. The sex of the egg(s) a female lays into a host are identifiable by observing wasp oviposition behaviour. Wasps exhibit different abdominal gestures when laying a female (fertilized) or male (unfertilized) egg. The sex of the polygerm is identified by staining with 2% acetolactic orcein and karyotyping (male = 11 (1n), female = 22 (2n)). For specific procedures, see refs 5 and 25. Recent studies indicate that some precocious larvae eclose in male broods at ecdysis to the host fifth instar¹⁹, but these larvae are present too late in the life cycle to have any influence on sibling rivalry.

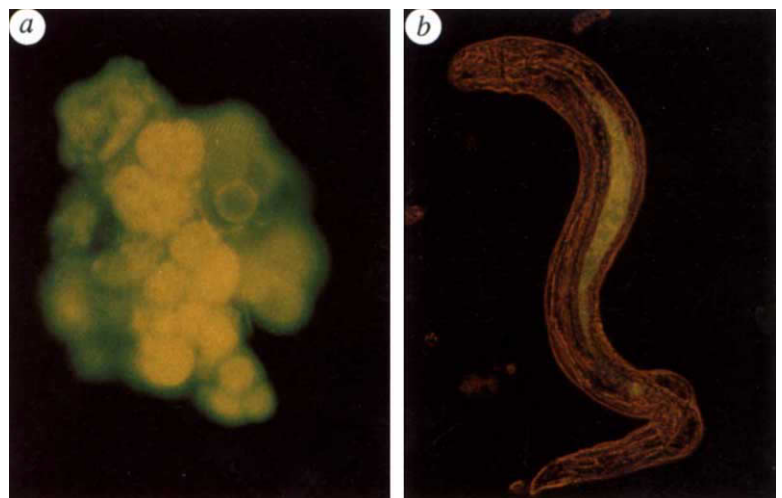
the host, a conflict could arise between brothers and sisters over the number of each sex that should be produced in a host. Such conflict is likely because: (1) all-female broods occur in the field and must mate with males from other hosts; (2) males are winged and disperse from mixed broods after mating with sisters; and (3) males live for 7–14 days^{5,6}.

This conflict appears to be resolved in favour of females through the siblicidal behaviour of precocious larvae. By killing brothers, precocious larvae secure more resources from the host for their clonal sisters. Selection may be stronger on females to reduce the number of males than on males to reduce the number of females because of relatedness asymmetries and different clutch-size-specific fitness functions. Brothers and sisters value their own sex equally ($r = 1$), but brothers are related to sisters by $1/2$, whereas sisters are related to brothers by $1/4$. Adult size (as estimated by tibial length) declines with increasing clutch size for both sexes, but female size decreases more rapidly than male size ($F_{2,212} = 15.6$, $P < 0.0001$). In turn, female fecundity is affected by wasp size more than male mating ability (P.J.O. and M.R.S., unpublished data). Given the opportunity that each clonal male in a mixed brood has to mate many sisters, a reduction in clutch size may benefit females more than males. The best male strategy may be to 'hide' in host tissue to avoid complete elimination by female precocious larvae.

Larval dimorphisms occur in other polyembryonic encyrtids and some aphids^{3,4,15,16}. It has been proposed that the 'fighter-morphs' in these clonal organisms defend reproductive siblings from predators and other parasites^{17–19}, but our results indicate that these larvae may play a larger role in sibling conflict than altruism. Parallels are found in parasitic wasps that lay a single egg per host and in which fighting larvae are common^{20–23,26}.

FIG. 2 a, Polygerm labelled with the vital tracer carboxy-fluorescein diacetate succinimidyl ester (CFSE)¹⁰ 24 h after injection into a host (158×). b, Presence of fluorescent tracer in the gut of a precocious larva 24 h after injection of a CFSE-labelled male polygerm into a host containing an all-female brood (80×).

METHODS. Polygerm (male or female) collected from fourth instar hosts was placed in physiological saline containing $1 \mu\text{mol}$ CFSE for 30 min. After rinsing (3×) in saline, individual polygerms were injected into a host containing a female brood of *C. floridanum*. Precocious larvae containing the tracer were identified by epifluorescent microscopy.



Siblicidal behaviour in a clonally reproducing species like *C. floridanum*, however, underscores the potential significance of sibling conflict in the evolution of reproductive strategies. □

Received 10 August; accepted 1 October 1992.

- Hamilton, W. D. *J. theor. Biol.* **7**, 1–52 (1964).
- Flanders, S. E. *Insects Soc.* **3**, 325–334 (1956).
- Silvestri, F. *Bull. Mus. comp. Zool. Harvard Univ.* **81**, 469–498 (1937).
- Ivanova-Kasas, O. M. in *Developmental Systems* (eds Counce, S. J. & Waddington, C. H.) 243–271 (Academic, New York, 1972).
- Strand, M. R. *J. Insect Behav.* **2**, 355–369 (1989).
- Strand, M. R. *Florida Entomol.* **72**, 32–42 (1989).
- Baehrecke, E. H., Grbic, M. & Strand, M. R. *J. exp. Zool.* **262**, 30–39 (1992).
- Baehrecke, E. H. & Strand, M. R. *J. Insect Morph. Embryol.* **19**, 165–175 (1990).
- Strand, M. R., Goodman, W. G. & Baehrecke, E. H. *Insect Biochem.* **21**, 205–214 (1991).
- Haugland, R. P. *Molecular Probes Handbook of Fluorescent Probes and Research Chemicals* 105–113 (Molecular Probes Inc., Eugene, Oregon, 1989).
- Taylor, P. D. *Nature* **291**, 64–66 (1981).
- Taylor, P. D. *Theoret. Pop. Biol.* **34**, 145–168 (1988).
- Grafen, A. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 62–84 (Blackwell, Oxford, 1984).
- Bulmer, M. G. *Heredity* **56**, 69–73 (1986).
- Aoki, S. *Kontyu* **47**, 390–398 (1977).
- Aoki, S. *Kontyu* **51**, 282–288 (1983).
- Cruz, Y. P. *Nature* **294**, 446–447 (1981).
- Aoki, S., Kurosu, U. & Usuba, S. *Kontyu* **52**, 458–460 (1984).
- Grbic, M. & Strand, M. R. in *Trichogramma and Other Egg Parasitoids* (eds Wajnberg, E. & Vinson, S. B.) 23–27 (INRA, Paris, 1990).
- Le Masurier, A. D. *Ecol. Ent.* **12**, 383–393 (1987).
- Suzuki, Y. & Iwasa, Y. *Res. Pop. Ecol.* **22**, 366–382 (1980).
- Godfray, H. C. J. *Am. Nat.* **129**, 221–233 (1987).
- Godfray, H. C. J. & Parker, G. A. *Phil. Trans. R. Soc. Lond.* **B332**, 67–79 (1991).
- Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, New York, 1981).
- Strand, M. R. & Ode, P. J. *Ann. Entomol. Soc. Am.* **83**, 834–837 (1990).
- Itô, Y. *Trends Ecol. Evol.* **4**, 69–73 (1989).

ACKNOWLEDGEMENTS. We thank E. Baehrecke, M. Louthier, J. Johnson, A. Ives and R. Jeanne for comments and assistance during the study. Financial support was provided by the USDA, Sigma Xi and the Fulbright Fellowship program.

Molecular and genetic damage in humans from environmental pollution in Poland

Frederica P. Perera*, Kari Hemminki†, Ewa Gryzbowska‡, Grazyna Motykiewicz‡, Jadwiga Michalska‡, Regina M. Santella*, Tie-Lan Young*, Christopher Dickey*, Paul Brandt-Rauf*, Immaculata DeVivo*, William Blaner§, Wei-Yann Tsai|| & Mieczyslaw Chorazy†

* Division of Environmental Sciences, Columbia University School of Public Health, 60 Haven Avenue, B1-109, New York, New York 10032, USA

† CNT, Karolinska Institute, Novum, S-14152 Huddinge, Sweden

‡ Department of Tumor Biology, Institute of Oncology, Gliwice P144-100, Poland

§ Columbia University, Institute of Human Nutrition, New York, New York 10032, USA

|| Columbia University, Division of Biostatistics, New York, New York 10032, USA

EXTREME environmental pollution such as that found in the highly industrialized Silesian region of Poland has been associated with increased risk of cancer and adverse reproductive outcomes^{1,2}. Among the most prevalent carcinogenic and mutagenic air pollutants in Silesia are the polycyclic aromatic hydrocarbons (PAH) which are largely produced by industrial and residential combustion of coal¹. Molecular epidemiology aims to prevent disease by using biological markers to identify risks well before clinical onset to allow effective intervention^{3–7}. Here, we use a battery of biological markers to measure molecular and genetic damage in peripheral blood samples from residents of Silesia and from persons living in a rural, less polluted area of Poland. The results show that their exposure to environmental pollution is associated

with significant increases in carcinogen–DNA adducts (PAH–DNA and aromatic adducts), in sister chromatid exchange including high-frequency cells, and in chromosomal aberrations as well as a doubling in the frequency of *ras* oncogene overexpression. We found that aromatic adducts on DNA were significantly correlated with chromosomal mutation, providing us with a molecular link between environmental exposure and a genetic alteration relevant to cancer and reproductive risk.

The environmentally exposed population is from the town of Gliwice in the Province of Katowice, Silesia, which is one of the most heavily polluted areas in the world and is characterized by high cancer mortality and high infant mortality¹. The average annual atmospheric concentration of benzo[*a*]pyrene, a representative PAH, as measured by ultraviolet spectrophotometry in Gliwice in 1987, was 66 ng m^{−3} (ref. 1). More recently, benzo[*a*]pyrene concentrations measured by gas chromatography and mass spectrometry were 57 ng m^{−3} in January and 15 ng m^{−3} in May⁸. This seasonal variation is attributable to the intensive combustion of coal for residential heating during the winter months and is paralleled by fluctuations in the mutagenic activity of air samples from this region^{9,10}.

A total of 88 males were enrolled in the study between January and October 1990, including 39 persons from Gliwice and 49 from Biala Podlaska, a rural province in northeastern Poland where winter atmospheric levels of benzo[*a*]pyrene are estimated to be 10-fold lower than in Silesia. A subset of subjects from Gliwice gave both winter and summer samples for adduct analysis. None of the Gliwice subjects was employed in PAH-generating industries; the controls had diverse non-farming

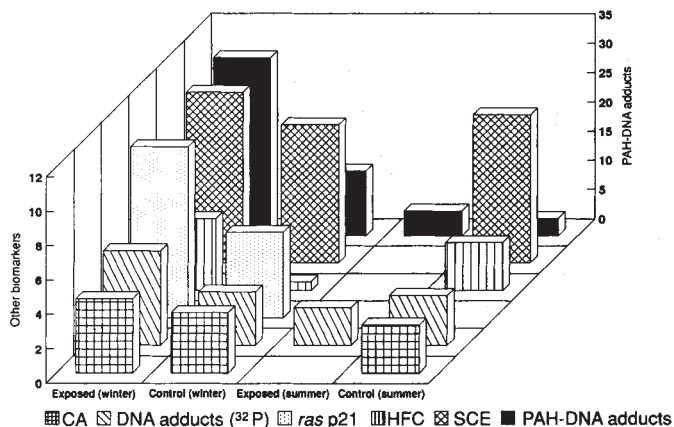


FIG. 1 Bar graph showing the mean levels of PAH–DNA adducts (by ELISA), SCE, *ras* p21 overexpression, DNA adducts (by ³²P-postlabelling) and CA for the four comparison groups. PAH–DNA adducts were analysed in coded samples by competitive ELISA, with fluorescence detection¹⁸, ³²P-postlabelling was also used to assay samples for aromatic adducts. Coded DNA samples (4 µg) were assayed in duplicate⁸. Quantitative differences between the two adduct assays are attributable to their differing specificity towards individual aromatic adducts. SCE, including HFC (the percentage of cells >2 standard deviations above the mean in terms of the number of exchanges per cell, here >17 SCE per cell), and CA were analysed as described^{19–21}, excluding individuals exposed to X-rays in the past 3 months^{18–20}. Plasma levels of *ras*-oncogene-related p21 proteins were determined by immunoblotting with monoclonal antibody¹¹. The p21 bands were densitometrically quantitated by volumetric integration of optical density using a laser densitometer and concentrations were determined by interpolation in comparison to volumetrically integrated optical densities of serial dilutions of known concentrations of recombinant human p21 protein run on the same blots. Elevated (positive) plasma levels of p21 were considered to be those concentrations in excess of 1.47 ng µl^{−1}, which is two standard deviations above the mean observed in plasma of 50 normal, healthy, non-smoking controls. Plasma retinol levels were determined by high-performance liquid chromatography^{22,23}. PAH–DNA and DNA adducts (³²P) are adducts per 10⁸ nucleotides; SCE, sister chromatid exchanges per cell; CA, chromosomal aberrations per 100 cells; HFC, high-frequency cells (percentage >17 SCE per cell); *ras* p21 overexpression (per cent positive).

TABLE 1 Results for six biological markers in seasonal comparisons of exposed and unexposed Polish residents

Markers	Exposed (winter)	Control (winter)	Exposed (summer)	Control (summer)	Intergroup comparisons <i>P</i> values*†
PAH-DNA (ELISA)					
adducts per 10 ⁸ nucleotides					EW > ES: <i>P</i> = 0.002
Mean (s.d.)	30.4 (31.7)	11.0 (22.6)	4.2 (13.5)	3.0 (3.3)	EW > CW: <i>P</i> = 0.001
<i>N</i>	20	21	21	22	ES > CS: <i>P</i> = 0.012
Range	1-73	1-73	1-63	1-13	CW > CS: <i>P</i> = 0.581
					CW > ES: <i>P</i> = 0.480
					EW > CS: <i>P</i> = 0.078
DNA adducts (³²P)					
adducts per 10 ⁸ nucleotides					EW > ES: <i>P</i> < 0.001
Mean (s.d.)	5.5 (4.5)	3.1 (1.7)	2.2 (3.3)	2.9 (2.8)	EW > CW: <i>P</i> = 0.066
<i>N</i>	32	16	22	22	ES < CS: <i>P</i> = 0.639
Range	1.2-18	0.52-7.5	0.3-16.3	0.6-11.1	CW > CS: <i>P</i> = 0.315
					CW > ES: <i>P</i> = 0.005
					EW > CS: <i>P</i> = 0.004
Sister chromatid exchange					
SCE per cell					
Mean (s.d.)	10 (1.5)	8.1 (1)	NA	8.7 (1.6)	EW > CW: <i>P</i> < 0.001
<i>N</i>	29	16		15	CW < CS: <i>P</i> = 0.353
Range	8-15	6-10		7-11	EW > CS: <i>P</i> = 0.014
High frequency cells					
per cent of cells counted					
Mean (s.d.)	4.2 (5.0)	0.5 (1.0)	NA	2.8 (3.0)	EW > CW: <i>P</i> < 0.001
<i>N</i>	30	16		15	CW < CS: <i>P</i> = 0.006
Range	0-22	0-4		0-10	EW > CS: <i>P</i> = 0.407
Chromosome aberrations					
CA per 100 cells					
Mean (s.d.)	4.3 (1.5)	3.5 (0.7)	NA	2.8 (0.8)	EW > CW: <i>P</i> = 0.392
<i>N</i>	15	2		16	CW > CS: <i>P</i> = 0.192
Range	2-8	3-4		2-5	EW > CS: <i>P</i> = 0.001
<i>ras</i> Onco protein (p21)					
Overexpression	4/39	1/22	NA	NA	EW to CW: <i>P</i> = 0.8
Per cent	10	5			

In cases where biomarker data were not normally distributed (adducts by ELISA and post-labelling), they were log-transformed in order to stabilize the variance before performing correlation analysis and ANOVA. The mean biomarker levels for each of the four groups (exposed, winter; control, winter; exposed, summer; control, summer) were compared by the Wilcoxon rank sum nonparametric test (2-tailed, $\alpha = 0.05$), except for adduct levels, which were compared using a modification of the Wilcoxon rank sum and signed ranks tests for combining paired and unpaired samples¹⁷. For CA, persons who were exposed to X-ray during the previous three months were excluded from the analysis. The proportions of individuals positive for *ras* oncoprotein overexpression in the exposed and control groups were compared by the Fisher's exact test. Next, analysis of variance (ANOVA) was used for intergroup comparisons adjusting for age and cigarette smoking. Finally, age, cigarette smoking and environmental exposure were used in a multiple regression model to estimate the linearity of the dose-response relationship for each of the biomarkers as well as interactions between these variables. Correlations between biomarkers were analysed by Pearson's product moment correlation analysis. NA, not available.

* EW: Exposed, winter; ES: exposed, summer; CW: control, winter; CS: control, summer.

† Modified Wilcoxon signed ranks and rank sum test or Wilcoxon rank sum test, except for *ras* p21 by Fisher's exact test.

occupations with no workplace exposure to PAHs. In interviews using a standardized questionnaire, subjects provided information on age, cigarette smoking, alcohol consumption, dietary and occupational exposure to PAH and exposure to X-rays. The Gliwice group had a mean age of 37 years (± 13) compared with controls (mean, 25 ± 7.5). The two groups were comparable in terms of their prevalence and frequency of smoking (70% smokers, mean 15 cigarettes per day).

Coded blood samples (30-50 ml) were collected in heparinized tubes in the winter and summer months as shown in table 1. Six different biological markers were measured in peripheral blood as shown in Table 1 and Fig. 1. They include the molecular dose or fingerprint of PAH in the form of carcinogen-DNA adducts (measured by enzyme-linked immunosorbent assay, ELISA), aromatic DNA adducts by ³²P-postlabelling (denoted as adducts (³²P)), and several different alterations in the structure or the expression of genes: sister chromatid exchanges (SCE), including high-frequency cells (HFC), chromosomal aberrations (CA) and activation of the *ras* oncogene. All of these biomarkers have been associated with lung cancer and/or occupational exposure to PAH^{3-7,11,12}. Plasma levels of retinol were analysed to see whether they influenced the levels of biomarkers because earlier studies have demonstrated that retinoids can inhibit PAH-DNA binding, mutagenesis and carcinogenesis¹³⁻¹⁵.

Table 1 shows the results of the assays, the numbers of samples analysed for each endpoint and the means ($\pm$ s.d.) and range for each of the four groups: those exposed during winter, those exposed during summer, and controls for the winter and summer periods. Sample from all four groups were analysed for PAH-DNA adducts and ³²P-postlabelled adducts. Sister chromatid exchange and chromosomal aberrations were not available for the exposed summer group. In some cases the amount or quality of sample was inadequate to carry out all six assays on the same individual.

Levels of PAH-DNA and aromatic DNA adducts (³²P) in the exposed winter group were significantly higher than in the corresponding summer samples ($P < 0.002$). PAH-DNA, SCE and HFC in the exposed winter samples were also significantly elevated compared with the control winter samples ($P < 0.001$). The proportion of positives for overexpression of the oncoprotein *ras* p21 was doubled in the exposed group (4/39; 10%) compared with controls (1/22; 5%), but the difference was not significant ($P = 0.8$). By analysis of variance (ANOVA), after adjusting for age and cigarette smoking, exposure significantly contributed to intergroup differences with respect to all biomarkers except HFC ($P < 0.05$). In the regression model, after adjusting for current smoking and age, exposure to environmental pollution was significantly related to PAH-DNA adducts

(by ELISA; $P=0.001$), to adducts (^{32}P) ($P=0.001$), to SCEs ($P=0.016$) and to CA including gaps ($P=0.01$). Possible interactions between smoking and environmental exposure were tested and proved to be nonsignificant and were therefore left out of the regression model. Retinol and biomarkers were not inversely associated. Regarding the subset of subjects from Gliwice who were repeatedly sampled, there was a significant difference between their winter and summer adduct levels by immunoassay (26.6 ± 32 ; $P=0.04$ versus 6.7 ± 18.6 , $n=11$) and by ^{32}P -postlabelling (6.1 ± 5.1 versus 2.3 ± 3.4 ; $P<0.001$, $n=21$). A number of biomarkers were intercorrelated: PAH-DNA adducts by ELISA with aromatic adducts by ^{32}P ($P=0.02$), adducts by ^{32}P with CA ($P=0.05$) and SCE with HFC ($P=0.001$). Other correlations were not significant. The range of biomarker values in the exposed winter group (Table 1) shows

that there was marked inter-individual variation in biomarkers: 73-fold for PAH-DNA (ELISA), 16-fold for DNA adducts (^{32}P), and 2–22-fold for SCE, HFC and CA. Probit analysis based on the PAH-DNA (ELISA) data in the most exposed group suggests that there would be very large variation (850-fold) in this biomarker among 95% of a standard population with comparable exposure.

Our study builds on earlier research in Poland⁸ and is the most comprehensive evaluation to date of multiple biomarkers in a population exposed to ambient environmental pollution. Although the ambient air concentrations of PAHs in Silesia are about 10–30-fold higher than in most Western countries^{8,16}, these results indicate that there is a non-threshold dose-response relationship for this class of pollutants. □

Received 20 July; accepted 12 October 1992.

- Hertzman, C. in *Health Policy Research Unit Vancouver, British Columbia* **90**, 2D (1990).
- Jedrychowski, W., Becher, H., Wahrendorf, J. & Basa-Cierpiele, Z. *J. A. Epidemiol. Commun. Health* **44**, 114–120 (1990).
- Perera, F. P. *CA—A Cancer J. Clin.* **40**, 227–228 (1990).
- Perera, F. P. & Schulte, P. A. (eds) *Molecular Epidemiology: Principles and Practices* (Academic, New York, in the press).
- Bartsch, H., Hemminki, K. & O'Neill, I. K. (eds) *Methods for Detecting DNA Damaging Agents in Humans*. International Agency for Research on Cancer, Lyon (1988).
- Hulka, B. S., Wilcosky, T. C. & Griffith, J. B. *Biological Markers in Epidemiology* (Oxford University Press, New York, 1990).
- Biomonitoring and Susceptibility Markers in Human Cancer: Applications in Molecular Epidemiology and Risk Assessment* (Monograph, International Agency for Research on Cancer, Lyon, in the press).
- Hemminki, K. *et al. Carcinogenesis* **11**, 1229–1231 (1990).
- Motykwicz, G., Szeliga, J., Cimander, B. & Chorazy, M. *Mutat. Res.* **223**, 243–251 (1989).
- Fukino, H., Mimura, S., Inoue, K. & Yamane, Y. *Mutat. Res.* **102**, 237–247 (1982).
- Brandt-Rauf, P. W. *et al. Ford Int. J. Cancer* **51**, 1–5 (1992).
- Brandt-Rauf, P. W., Smith, S. J. & Perera, F. P. in *Current Issues in Respiratory Public Health* (eds Witek, T. J. & Schacter, E. N.) Vol. 3 44–61 (Lippincott, New York, 1990).

- Moon, T. E. & Micozzi, M. S. (eds) *Nutrition and Cancer Prevention: Investigating the Role of Macronutrients* (Dekker, New York, 1991).
- Hennekens, C. H., Stampfer, M. J. & Willett, W. *Cancer Detect. Prev.* **7**, 147–158 (1984).
- Pezzuto, J. M., Chou, J., Jia-Xiong, J., Kudo, S. & Swanson, S. M. *Proc. A. Mtg Amer. Ass. Cancer Res.* **27**, 126 (1986).
- Faoro, R. *Aerometric Information Retrieval System (AIRS) Data for 1985–1990* (Environmental Protection Agency, Research Triangle Park, North Carolina (1991).
- Wei, L. J. *J. Am. Stat. Assoc.* **78**, 725–729 (1983).
- Perera, F. P. *et al. Cancer Res.* **48**, 2288–2291 (1988).
- Carrano, A. V. & Natarajan, A. T. *Mutat. Res.* **204**, 379–407 (1988).
- Moore, D. H. & Carrano, A. V. in *Sister Chromatid Exchanges* (eds Tice, R. R. & Hollaender, A. 469–480 (Plenum, New York, 1984).
- Mayer, J. L. *et al. Mutation Res.* (in the press).
- Bieri, J. G., Tolliver, T. J. & Catignani, G. L. *Am. J. Clin. Nutr.* **32**, 2500–2507 (1979).
- Friedman, G. D. *et al. Am. J. Epidemiol.* **123**, 781–789 (1986).

ACKNOWLEDGEMENTS. We thank I. B. Weinstein, D. Tang, D. Hattis, A. Kostowska, M. Krawczyk, H. Paterak and R. Piantadosi for help with data and laboratory analysis. This work was supported by National Cancer Institute, the National Institute of Occupational Safety and Health, the National Institute of Environmental Health Sciences, and the United Nations Development Program.

GM-CSF and TNF- α cooperate in the generation of dendritic Langerhans cells

C. Caux*, C. Dezutter-Dambuyant†, D. Schmitt† & J. Banchereau*

* Schering-Plough, Laboratory for Immunological Research, 27 chemin des Peupliers, BP 11, 69571 Dardilly, France

† INSERM U346, Hôpital E. Herriot, 5 place d'Arsonvae, 69437 Lyon, Cedex 03, France

DENDRITIC cells comprise a system of highly efficient antigen-presenting cells which initiate immune responses such as the sensitization of T cells restricted by major histocompatibility complex molecules, the rejection of organ transplants and the formation of T-cell-dependent antibodies. Dendritic cells are found in many non-lymphoid tissues, such as skin (Langerhans cells) and mucosa, and they migrate after antigen capture through the afferent lymph or the bloodstream to lymphoid organs, where they efficiently present antigen to T cells¹. Dendritic cells are difficult to isolate and, although they originate from bone marrow^{2,3} their site of maturation and the conditions that direct their growth and differentiation are still poorly characterized. Granulocyte macrophage-colony stimulating factor (GM-CSF) favours the outgrowth of dendritic cells from mouse peripheral blood⁴. Here we extend this finding to man and demonstrate that cooperation between GM-CSF and tumour necrosis factor- α (TNF- α) is crucial for the generation of human dendritic/Langerhans cells from CD34⁺ haematopoietic progenitors. The availability of large numbers of these cells should now facilitate the understanding of their role in immunological regulation and disorder.

We have shown that TNF- α stimulates interleukin-3 and GM-CSF-dependent early myelopoiesis *in vitro* (ref. 5, and

unpublished work) and facilitates the development of monocytic cells from CD34⁺ haematopoietic progenitor cells (HPC)⁶. Culturing cord blood CD34⁺ HPC for 12 days with GM-CSF $\pm$ TNF- α results in the generation of adherent and non-adherent cells, whereas TNF- α alone has no effect on HPC development⁵. In the absence of TNF- α , both cell types appear as regularly shaped monocytes, macrophages and granulocytes (Figs 1a, c and 2a). In contrast, adherent cells and 20–50% of non-adherent cells obtained with GM-CSF + TNF- α display a dendritic morphology with delicate membrane projections (Figs 1b, d and 2b). Freshly isolated HPC do not bear CD1a (a molecule specifically expressed on Langerhans cells⁷ and thymocytes), and culturing them for 12 days with GM-CSF results in the generation of 5–15% CD1a⁺ cells (Fig. 3b). Adding TNF- α and GM-CSF to HPC cultures increases the total cell number (Fig. 3a) and the proportion of CD1a⁺ cells (Fig. 3b), resulting in 10–20-fold increase of viable CD1a cells. Thus it is possible to obtain from $0.5\text{--}1 \times 10^6$ HPC as many as $1\text{--}2.5 \times 10^7$ viable CD1a⁺ cells within two weeks. Examination by electron microscopy shows that HPC-derived CD1a⁺, HLA-DR⁺ cells have a lobulated nucleus and a villous surface with dendritic projections (Fig. 1e, f). After immunoprecipitation they have a typical CD1a molecule of M_r 49,000. Furthermore, one out of five CD1a⁺ cells contained Birbeck granules (Fig. 1g–i), organelles with double-membrane junctions that are characteristic of Langerhans cells. CD1a⁺ cells express antigens (such as CD4, CD40, B7/BB1 and HLA class II) previously identified on dendritic cells isolated from either tissues or blood (Table 1).

As dendritic cells are potent stimulators of allogeneic T cells^{8,9}, we assayed HPC cultured for 12 days with GM-CSF $\pm$ TNF- α for their capacity to induce proliferation of resting allogeneic DC4⁺ T cells. As shown in Fig. 3c, cells cultured with GM-CSF alone or with GM-CSF + TNF- α induced a strong proliferation of allogeneic CD4⁺ T cells. Tritiated thymidine incorporation of 2.5×10^4 CD4⁺ T cells is increased 50-fold by 6,500 (1,000–8,000)

FIG. 1 Morphology of cells generated from CD34⁺ HPC in response to GM-CSF and GM-CSF+TNF- α . CD34⁺HPC were cultured for 12 days with GM-CSF alone (*a* and *c*) or GM-CSF+TNF- α (*b* and *d-i*). *a* and *b*, Adherent cells (scale bar, 10 μ m); *c* and *d*, non-adherent cells (scale bar, 10 μ m) examined under phase-contrast microscopy. *e-i*, Transmission electron microscopographs; *e*, CD1a (10 nm), immunogold labelling (scale bar, 1 μ m); *f*, CD1a (5 nm) and HLA-DR (17 nm), immunogold double-labelling (scale bar, 100 nm); *g*, portion of *e* at higher magnification: arrows indicate Birbeck granules (scale bar, 1 μ m); *h* and *i*, Birbeck granules from *e* (scale bar, 100 nm).

METHODS. CD34⁺ HPC isolation and culture: CD34⁺ HPC (95 to 99% pure) were isolated from cord blood cells through positive selection by indirect immune 'panning' and immunomagnetic mature cell depletion⁵. CD34⁺ HPC (1×10^5 per ml) were cultured with 100 ng ml⁻¹ (200 U ml⁻¹) recombinant human GM-CSF (2×10^6 U mg⁻¹; Schering) $\pm$ 2.5 ng ml⁻¹ (50 U ml⁻¹) recombinant purified human TNF- α (2×10^7 U mg⁻¹; Genzyme) in RPMI 1640 10% fetal bovine serum (FBS)⁵. Cultures were split every 4-5 days. Indirect immunogold-labelling and transmission electron microscopy preparations: Cells were incubated with DMC1 mAb (anti-CD1a)²⁸ (for 1 h at 4 $^{\circ}$ C), gold-particle-conjugated anti-mouse immunoglobulin (10 nm, single staining; 5 nm, double staining; Amersham) (for 1 h at 4 $^{\circ}$ C), and mouse serum (for 30 min at 4 $^{\circ}$ C) and HLA-DR BL-2 (17 nm; Immunotech) (for 1 h at 4 $^{\circ}$ C) for double staining. After centrifugation, the pellet was fixed and processed for transmission electron microscopy as described²⁹. Ultra-thin sections were stained with lead citrate and uranyl acetate and examined under a Jeol 1200 EX electron microscope at 80 kV accelerating voltage.

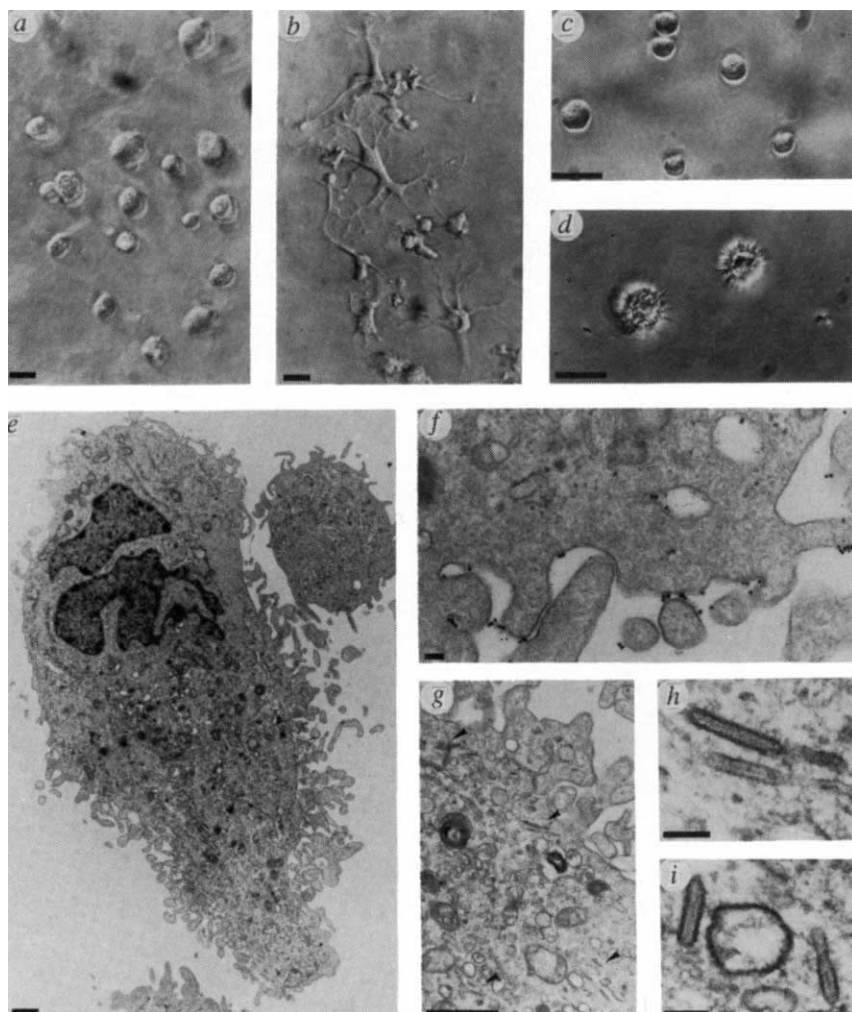
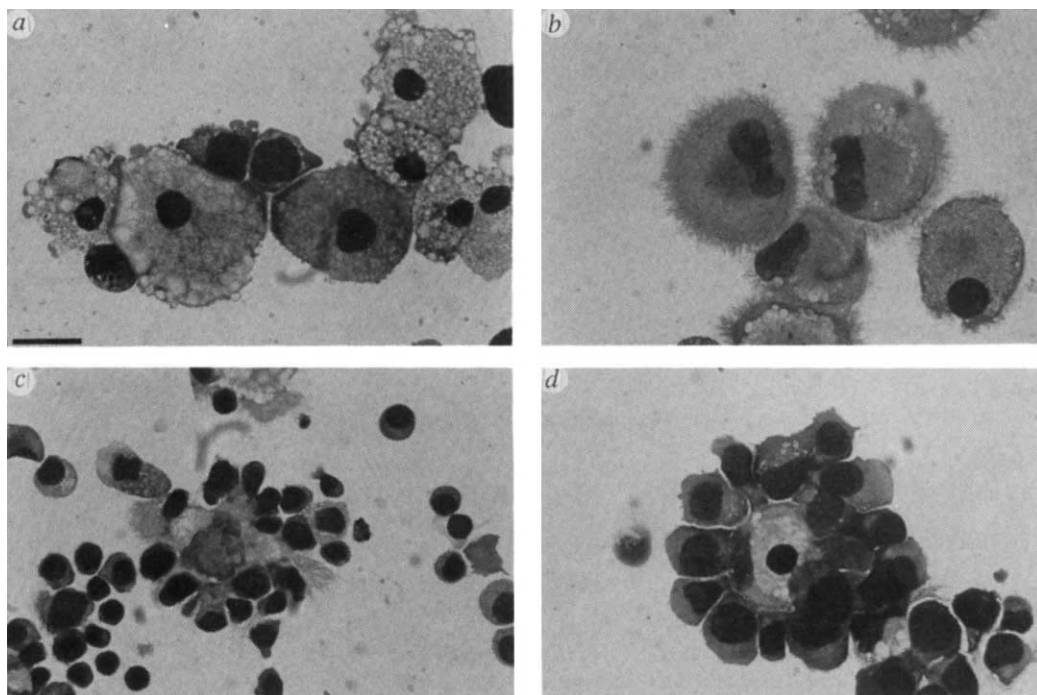


FIG. 2 Morphology in May Grünwald Giemsa staining of CD34⁺ HPC-derived cells, alone and interacting with allogeneic CD4⁺T cells. *a* and *b*, Cells generated in response to GM-CSF alone and GM-CSF+TNF- α , respectively (scale bar, 10 μ m); *c* and *d*, GM-CSF+TNF- α generated cells after respectively 24 h and 96 h co-culture with allogeneic CD4⁺ T cells (scale bar, 10 μ m).

METHODS. CD4⁺ T lymphocyte purification: CD4⁺ T lymphocytes were purified from non-adherent mononuclear peripheral blood cells by immunomagnetic depletion using a cocktail of mAbs (against CD14, CD16, CD20, CD8, CD40). After 2 rounds of bead depletion, purity of CD4⁺ cells was routinely higher than 95%. 10^5 GM-CSF+TNF- α cells were cultured with 10^6 allogeneic resting CD4⁺ T cells in 1 ml RPMI 1640 supplemented with 10% human AB⁺ serum. After 24 h and 96 h, cells were stained with May Grünwald Giemsa.



and 60 (60–250) cells (means and ranges from five experiments) from GM-CSF and GM-CSF+TNF- α respectively. Allogeneic CD4⁺ T cells derived from either cord or adult blood were equally stimulated, whereas syngeneic CD4⁺ T cells were only weakly, although significantly, stimulated (not shown). As few as ten CD1a⁺ cells induced a strong proliferation of allogeneic CD4⁺ T cells, but CD1a⁻ cells were poor antigen-presenting cells (Fig. 3d). Allogeneic CD4⁺ T cells form rosettes with GM-CSF+TNF- α cells within 24 h (Fig. 2c) and take on a blast morphology after 96 h of co-culture (Fig. 2d). Thus, cells cultured with GM-CSF+TNF- α have a high antigen-presenting capacity to T cells.

This study indicates that the combination of GM-CSF and TNF- α induces the differentiation of human HPC purified from cord blood or bone marrow (results not shown) into cells displaying the morphology, phenotype and function of dendritic cells and, more particularly, Langerhans cells. Interestingly, GM-CSF has previously been recognized as being able to stimulate expression of class II major histocompatibility complex and

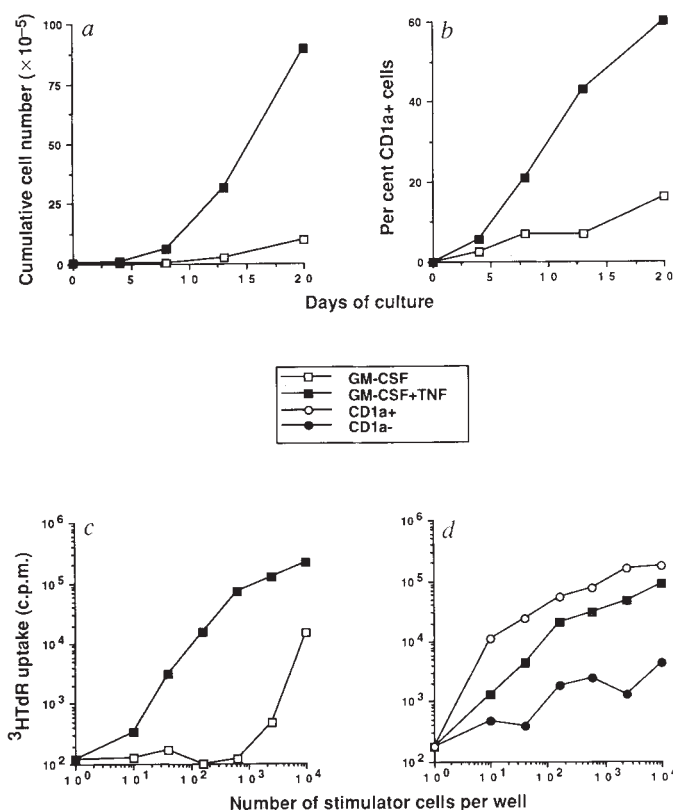


FIG. 3 Kinetics of CD1a⁺ cell generation and stimulation of primary allogeneic CD4⁺ T-cell proliferation. 10^5 CD34⁺ cord blood HPC were expanded in RPMI 1640 10% FBS with GM-CSF or GM-CSF+TNF- α . Cell numbers (a) and CD1a expression (b) were monitored. Cells were cultured for 12 days in the presence of GM-CSF and GM-CSF+TNF- α (c), and for some experiments GM-CSF+TNF- α cells were FACS-sorted according to CD1a expression using a FACStar (Becton-Dickinson) (d) and used as stimulator for adult allogeneic CD4⁺ T-cell proliferation data are from one representative experiment.

METHODS. Indirect immunofluorescence was done according to standard techniques⁶. Fluorescence was analysed with a FACScan flow-cytometer (Becton-Dickinson). Allogeneic CD4⁺ T cell proliferation: After 12 days of culture in RPMI 1640 and 10% FBS, cells generated from CD34⁺ cord blood HPC were irradiated (4,000 rads) and used as stimulator cells for resting allogeneic CD4⁺ T cells. Between 10 and 1×10^4 stimulator cells were seeded for 2.5×10^4 resting CD4⁺ T cells in round-bottomed microtest tissue-culture plates in RPMI 1640 with 10% human AB⁺ serum. After 5 days incubation, cells were pulsed with 1 μ Ci or 3 H-thymidine (3 H]TdR) for 6 h collected and counted. Tests were carried out in triplicate; results are expressed as mean counts per minute (c.p.m.). Standard deviations always represent less than 10% of the mean value.

TABLE 1 Comparison of phenotype of cells cultured with GM-CSF or GM-CSF+TNF- α with that of skin Langerhans cells and blood dendritic cells

Surface antigens	Cultured cells		Freshly isolated cells ^{1,20,27}	
	GM-CSF +TNF- α CD1a ⁺ cells	GM-CSF CD14 ⁺ cells	Langerhans cells	Dendritic cells
CD1a	+	—	+	—
CD1b	—	—	—	+/-3
CD1c	+	—	+	—
CD14	+/-	+	+/-	—
CD4	+	+/-	+	+/-
CD40	+	+/-	+	+
B7/BB1	+	—	ND	+
CD16 (Fc γ RIII)	+/-	+	—	—
CD32 (Fc γ RII)	+/-	+	—	—
CD64 (Fc γ RI)	—	+	—	—
CD21 (CR2)	—	—	—	—
CD11b (CR3)	+	+	+	—
CD35 (CR1)	—	+	—	—
CD54 (ICAM)	+	+	+	+
CD11a (LFA 1 α)	+	+	+	+
CD18 (LFA- β)	+	+	+	+
HLA-DR*	+	+	+	+
HLA-DQ*	+	+	+	+

Double-colour fluorescence of cord blood CD34⁺HPC cultured for 12 d in the presence of GM-CSF or GM-CSF+TNF- α was carried out by sequential incubation of the cells with unconjugated monoclonal antibodies (mAbs), phycoerythrin conjugated anti-mouse immunoglobulin, normal mouse serum and FITC-OKT6 (CD1a; Ortho) or FITC-LeuM3 (CD14; Becton-Dickinson) mAbs. 20–50% of cells in GM-CSF+TNF- α cultures express CD1a (and represent dendritic cells) and 20–50% of cells in GM-CSF cultures express CD14 (and represent monocytes/macrophages). Other cells in both culture conditions are undifferentiated blasts and immature granulocytes. Negative controls were performed with unrelated murine mAbs (not shown). Fluorescence analysis was determined with a FACScan flow cytometer. A plus sign means that >90% of cells were positive; a minus sign indicates that <10% of cells were positive; +/- means that between 10 and 90% of cells were positive. ND, not determined.

* GM-CSF+TNF- α CD1a⁺ cells express 4 to 10 times more HLA class II molecules than GM-CSF CD14⁺ cells.

antigen-presenting capacity in Langerhans cells^{10,11}, and TNF- α was found to maintain the viability of cultured Langerhans cells¹². Thus, GM-CSF and TNF- α act both on the differentiation of dendritic precursor cells and the activation of mature dendritic cells. The rarity of dendritic cell precursors among bone marrow cells suggests that dendritic cells may mature outside bone marrow. Thus, circulating CD34⁺HPC (ref. 13) could generate a localized extramedullary haematopoiesis, yielding professional antigen-presenting cells following the release of TNF- α and GM-CSF by non-antigen-specific cells such as monocytes, neutrophils and keratinocytes^{14–16} in response to pathogen injury. This would represent a bridge between non-antigen-specific defence mechanisms and an antigen-specific adaptive immune system.

This influence of TNF- α on the generation of dendritic cells may explain its antitumour effect *in vivo*¹⁷; this may derive in part from an enhanced generation of dendritic cells and an increased generation of tumour-specific cytotoxic T cells. It may also explain the stimulatory effects of a single injection of TNF- α on the production of antibody to T-cell-dependent antigens¹⁸.

The production of 'naive' dendritic cells in such large numbers, compared to the yield from standard isolation procedures from blood^{19,20} or skin²¹, should help achieve a better understanding of the physiology of these cells, which may be important for example in human immunodeficiency virus infection and related diseases^{22–25}. Their strong antigen-presenting capacity could facilitate the study of primary T cell and B cell antigen-specific responses. Furthermore, dendritic cells generated *in vitro* could be used after antigen pulse²⁶ for immunization, and the

combined administration of GM-CSF and TNF- α might boost immune responses *in vivo*. □

Received 20 May; accepted 25 September 1992.

- Steinman, R. M. *A. Rev. Immun.* **9**, 271–296 (1991).
- Steinman, R. M., Lustig, D. S. & Cohn, Z. A. *J. exp. Med.* **139**, 1431–1445 (1974).
- Katz, S. I., Tamaki, K. & Sachs, D. H. *Nature* **282**, 324–326 (1979).
- Inaba, K. *et al. J. exp. Med.* **175**, 1157–1167 (1992).
- Caux, C. *et al. Blood* **75**, 2292–2298 (1990).
- Caux, C. *et al. Blood* **78**, 635–644 (1991).
- van de Rijn, M. *et al. Hum. Immun.* **9**, 201–210 (1984).
- Van Voorhis, W. C. *et al. J. exp. Med.* **158**, 171–191 (1983).
- Crow, M. K. & Kunkel, H. G. *Clin. exp. Immun.* **49**, 338–346 (1982).
- Witmer-Pack, M. D., Olivier, W., Valinsky, J., Schuler, G. & Steinman, R. M. *J. exp. Med.* **166**, 1484–1498 (1987).
- Heufler, C., Koch, F. & Schuler, G. *J. exp. Med.* **167**, 700–705 (1987).
- Koch, F. *et al. J. exp. Med.* **171**, 159–172 (1990).
- Gabbianelli, M. *et al. Science* **249**, 1561–1564 (1990).
- Djeu, J. Y. in *Tumour Necrosis Factors: The Molecules and Their Emerging Role in Medicine* (ed. Beutler, B.) 531–537 (Raven, New York, 1992).
- Metcalf, D. *Science* **254**, 529–533 (1991).
- Luger, T. A. *Acta derm. venerol.* **69**, 61–76 (1989).
- Malik, T. A. in *Tumour Necrosis Factors: The Molecules and Their Emerging Role in Medicine* (ed. Beutler, B.) 407–423 (Raven, New York, 1992).
- Ghiara, P., Boraschi, D., Neucroni, L., Ghezzi, P. & Togliabue, A. *J. Immun.* **139**, 1753–1758 (1987).
- Van Voorhis, W. C., Hair, L. S., Steinman, R. M. & Kaplan, G. *J. exp. Med.* **155**, 1172–1187 (1982).
- Freudenthal, P. S. & Steinman, R. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7698–7702 (1990).
- Romani, N. *et al. J. invest. Dermatol.* **93**, 600–609 (1989).
- Knight, S. C., Macatonia, S. E., Bedford, P. A. & Patterson, S. in *Accessory Cells in HIV and Other Retroviral Infections* (eds Racz, P., Dijkstra, C. D. & Gluckman, J. C.) 145–154 (Karger, Basel, 1991).
- Langhoff, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 7998–8002 (1991).
- Tschachler, E. *et al. J. invest. Dermatol.* **88**, 233–237 (1987).
- Cameron, P. U. *et al. Science* **257**, 383–387 (1992).
- Sornasse, T. *et al. J. exp. Med.* **175**, 15–21 (1992).
- Dezutter-Dambuyant, C. *et al. Hybridoma* **8**, 199–208 (1989).
- Dezutter-Dambuyant, C., Schmikt, D., Faure, M., Horisberger, M. & Thivolet, J. *J. invest. Derm.* **84**, 465–468 (1985).
- Hart, D. N. J. & McKenzie, J. L. *J. exp. Med.* **168**, 157–160 (1988).

ACKNOWLEDGEMENTS. We thank B. Vanberviet, V. Duvert, and I. Durand for technical assistance; N. Courbière and M. Vatan for editorial help; J. P. Favier, R. C. Rudigoz, J. M. Thoulon and colleagues for umbilical cord blood samples; and J. Chiffler, J. A. Waitz, T. Mosmann and S. Saeland for support and for critical reading of the manuscript.

Enhancing protein C interaction with thrombin results in a clot-activated anticoagulant

Mark A. Richardson, Bruce Gerlitz & Brian W. Grinnell*

Department of Cardiovascular Research, Lilly Research Laboratories, Lilly Corporate Center, Indianapolis, Indiana 46285-0424, USA

HUMAN protein C is a vitamin K-dependent plasma glycoprotein that circulates as an inactive zymogen. At the endothelial cell surface, thrombin in complex with the integral membrane protein thrombomodulin converts protein C to its active form by specific cleavage of an activation peptide^{1–3}. The activated form of protein C has potent anticoagulant activity as a feedback regulator of thrombin generation (reviewed in refs 4–6), and also has profibrinolytic^{7–10}, anti-ischaemic¹¹ and anti-inflammatory properties¹². Protein C is effective in the treatment of model and human thrombotic diseases^{4,13–16} but, except when it has been used to treat genetic or acquired deficiencies and microvascular thrombosis, it is administered as the activated enzyme, which has a short biological half-life. We have altered two putative inhibitory acidic residues near the thrombin cleavage site, which results in a 30-fold increase in substrate utilization by α -thrombin. We combined these changes with a genetically altered glycoform to generate a zymogen protein C with a 60-fold increased cleavage rate by free α -thrombin, independent of its cofactor thrombomodulin. We show that this 'proform' of protein C, unlike the natural circulating zymogen, can be activated by thrombin generated in clotting human plasma, resulting in an inhibition of further clot formation. Our data

therefore show that we have engineered a site-activated agent, which only has anticoagulant activity when significant amounts of thrombin are being generated.

The native human protein C molecule (Fig. 1a) circulates predominantly as a disulphide-linked heterodimer^{4,17} composed of a light chain (M_r 25 K) and a heavy chain (M_r 41K) containing the serine protease domain, which is preceded by a 12-amino-acid N-terminal activation peptide. Under physiological conditions the conversion of zymogen protein C to activated protein C by thrombin alone is inhibited by calcium, unless it is associated with the endothelial cofactor thrombomodulin. Thus, administration of the zymogen may not be effective for the treatment of macrovascular thrombosis where surface thrombomodulin is limiting¹⁸.

The amino acids involved in the efficient recognition and cleavage of thrombin substrates have been studied^{19,20}. For example, the cleavage by thrombin of eel calcitonin with an aspartate residue at P2' is inefficient when compared with salmon calcitonin with an asparagine at this site¹⁹. Similarly, there are acidic residues surrounding the thrombin cleavage site in human protein C at both the P3 and P3' positions, and it has been suggested that these residues may inhibit thrombin cleavage, particularly in the absence of thrombomodulin^{21–24}. We previously constructed a derivative of human protein C, D167F, which resulted in an increase in thrombin activation due to a reduction in the inhibitory effect of calcium²⁴.

Here we have eliminated the P3' acidic residue in human protein C by point mutation or by substitution of the cleavage sequences from salmon calcitonin for the activation peptide region. We first substituted asparagine at the P2' position, which is analogous to that in the efficiently cleaved salmon calcitonin sequence, and we deleted the acidic residue at P3'. These changes, as well as substitution of the salmon calcitonin sequence for the activation peptide, resulted in a 400 and 200-fold increase in the rate of cleavage by free α -thrombin, respectively. These alterations resulted, however, in a ~95% reduction in protein C functional protease activity. Although the presence of asparagine at the P2' position gave high rates of thrombin activation, mutagenesis showed there was a strict requirement for the conserved hydrophobic residue at P2' for full activity (data not shown).

We have, however, defined specific single amino-acid changes, including alterations in the glycosylation pattern²⁵, that result in protein C molecules with an apparent increased substrate selectivity for thrombin and in the maintenance or enhancement of their activity. As shown in Fig. 1, single-point alterations in protein C have been defined that result in modest increases in the rate of thrombin activation in the presence of calcium: for example N313Q and D172N result in 2- and 4-fold increases, respectively. The mutant D172N demonstrates that the presence of the P3' acidic residue inhibits thrombin cleavage. The effect of neutralizing the acidic residues at P3 and P3' gave a multiplicative effect of ~30-fold, confirming the inhibitory role of these two acidic residues near the thrombin cleavage site. The modest individual rate effects of the single point mutations, when combined as in derivative FLIN-Q3, resulted in 60 times the apparent substrate selectivity of the wild-type protein. Under our assay conditions, the stimulatory effect of cofactor thrombomodulin on the thrombin cleavage rate of wild-type human protein C is ~380-fold. Thus, elimination of both acidic residues at P3 and P3', together with the glycosylation site change in N313Q, gives a derivative with a reduced requirement for thrombomodulin cofactor activity.

Table 1 shows the anticoagulant activities of the derivatives in Fig. 1 after activation by thrombin. Deglycosylation of the heavy chain increases the activity of the protein, presumably by reducing net charge²⁵, as seen for both the N313Q and FLIN-Q3 derivatives (Table 1), which have the glycosylation site at amino-acid residue 313 eliminated. The changes in D167F and D172N did not affect their anticoagulant activity, which was normal.

* To whom correspondence should be addressed.

TABLE 1 Functional anticoagulant activity of human protein C derivatives

Protein C	Anticoagulant activity (units per mg)	Relative activity
Wild type	325 ± 65	1
N313Q	577 ± 17	1.8
D172N	289 ± 13	0.9
D167F	283 ± 27	0.9
FLIN	313 ± 65	1
FLIN-Q3	552 ± 37	1.7

Fully activated protein C was obtained by activating material with 10 nM thrombin in complex with soluble recombinant human thrombomodulin TMD1-75 (ref. 30). The anticoagulant activity of the activated molecules was measured with an activated partial thromboplastin time clotting assay. One unit of anticoagulant activity is defined as the equivalent of 4 µg plasma protein C, which is normally present in 1 ml plasma³¹. Thus, fully activated plasma-derived protein C has a defined specific activity of 250 units mg⁻¹. The active site of the molecules was necessary for anticoagulant activity, as activity could be completely eliminated by titration with dansyl-Glu-Gly-Arg-chloromethylketone, an irreversible active-site reagent.

The FLIN-Q3 derivative was also activated more efficiently than wild-type protein C by thrombin in complex with thrombomodulin (Fig. 2), with the initial rate of activation being 9.2 ± 1.0 times higher. Thrombomodulin alters the substrate specificity of thrombin through a conformational change²⁶, which may overcome the inhibitory effect of acidic residues in protein C. Eliminating the acidic residues at P3 and P3' gave only a 3-fold increase in the rate of activation by the thrombin-thrombomodulin complex (data not shown), compared with the ~30-fold effect on the rate by thrombin alone (FLIN in Fig. 1.). Thus, the changes in protein C that affect thrombin rates of

FIG. 1 Schematic representation of protein C single-chain precursor with *a*, the location of site-specific amino-acid changes and *b*, the effect of these changes on the rate of thrombin activation in the presence of 3 mM calcium. Numbers refer to amino-acid positions in the mature protein (after removal of the signal sequence and propeptide). The site for thrombin cleavage of the activation peptide (AP) is indicated. The designations P3 and P3' refer to the third position N-terminal and C-terminal, respectively, to the thrombin cleavage site. The first letter preceding the amino-acid number is the one-letter code for the natural amino acid; the letter after the number refers to the change by mutagenesis. The light chain contains a region of γ -carboxyglutamic acid (GLA) residues, which is necessary for calcium-dependent membrane binding and functional activity (reviewed in ref. 5), and two domains homologous to epidermal growth factor (EGF).

METHODS. Each amino-acid change was made by site-directed mutagenesis using the cDNA coding sequence³². A 730-base-pair *SalI/SacI* fragment (corresponding to a region from just N-terminal of the first EGF domain through to roughly half of the serine protease domain) was isolated from the plasmid pLPC (ref. 31) and cloned into M13mp18 and into M13mp19. These clones, designated MP18-nHPC.730 and MP19-nHPC.730, were used as the source of single-stranded DNA template for generation of mutants. Details of the mutagenesis have been described²⁵, as have the primers used for N313Q²⁵ and D167F²⁴. Primers for the other amino-acid changes were: D172N: CCAAGAAGACCAAGTAGATCCAAGGCTCATTAATGGGAAGATGACCAAGGC; FLIN: GACCAAGAAGACCAAGTATTTCCAAGGCTCATTAATGGGAAGATGACCAAGGC; FLIN-Q3 was constructed by restriction-fragment ligation of the FLIN and N313Q coding sequences. The University of Wisconsin GCC sequence analysis software package version 7.0 was used for oligonucleotide and restriction site analyses³³. Construction of vectors for protein C expression, the development of recombinant cell lines, and the purification of each derivative from conditioned culture medium were all as described^{25,31}. Activation rates were determined at 37 °C using human thrombin (10 nM) in a reaction mix containing 20 mM Tris, pH 7.4, 150 mM NaCl, 0.1 mg ml⁻¹ BSA, 3 mM CaCl₂. Purified protein C, both wild-type and derivatives, were used at a concentration of from 0.81 to 1.61 µM in the activation reaction. Activation rates were determined by removing aliquots from the reaction at various time points to a 96-well plate, adding the chromogenic substrate (S-2366; Chromogenix) to a final concentration of 0.75 mM, and measuring the change in absorbance per min at 405 nm in a ThermoMax kinetic microtitre plate reader (Molecular Devices). Rates were determined by converting change in absorbance per minute to the amount

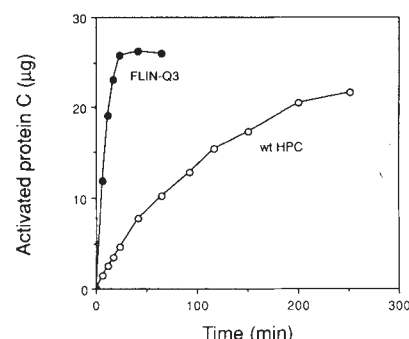
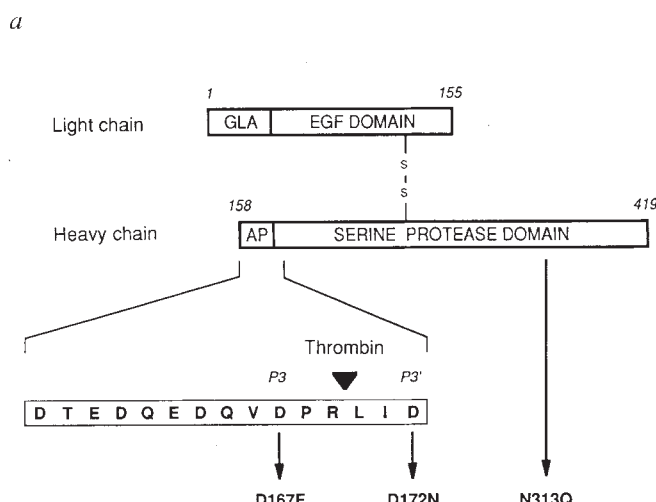


FIG. 2 Activation of wild-type human protein C (wt HPC) and FLIN-Q3 by thrombin complexed with thrombomodulin. Each protein C was activated at 10 nM thrombin and at a 15-fold molar excess of human recombinant soluble thrombomodulin at 37 °C (ref. 30). The amount of the activated form of protein C was determined as described in the legend to Fig. 1.

activation are not generalized alterations because most of the stimulation resulting from neutralizing the acidic residues is eliminated by the conformational change induced by thrombomodulin. There are comparable differences in the rate of wild-type human protein C activation with and without thrombomodulin by the thrombin mutant E192Q, which mimics the catalytic switch induced by thrombomodulin²¹. The ~10-fold increase in FLIN-Q3 activation rate with the thrombin-thrombomodulin complex was primarily due to the 3-fold increase in rate from elimination of the P3 and P3' acidic residues, combined with a 3-fold effect at N313Q. This effect was mainly attributable



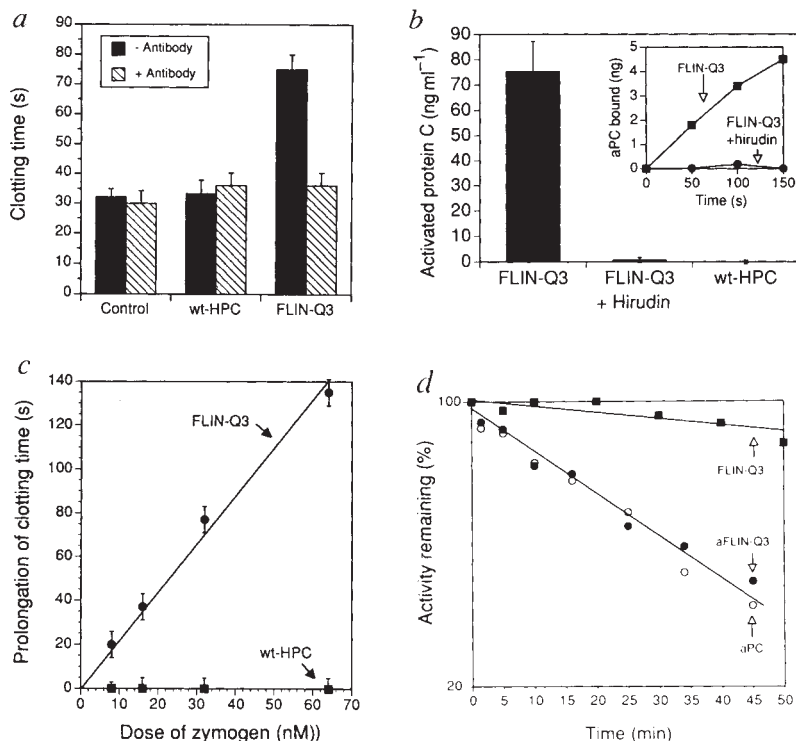
b

Protein C	Amino-acid changes	Thrombin activation rate (ng min ⁻¹)	Thrombin activation relative rates
wt-HPC	none	0.5 ± 0.1	1
N313Q	Asn 313 to Gln	1.1 ± 0.2	2
D172N	Asp 172 to Asn	2.2 ± 0.4	4
D167F	Asp 167 to Phe	6.3 ± 1.0	12
FLIN	Asp 167 to Phe Asp 172 to Asn	15.9 ± 4.0	30
FLIN-Q3	Asp 167 to Phe Asp 172 to Asn Asn 313 to Gln	32.3 ± 6.4	61

of activated protein C generated, using the specific activities determined for each protein, and plotted versus activation time. The amount of activated protein C produced was less than 10% of the initial zymogen in all experiments.

FIG. 3 Activity of zymogen FLIN-Q3 in clotting human plasma. **a**, Comparison of clotting times in human plasma in the presence of wild-type human protein C (wt HPC) or FLIN-Q3, and in the presence or absence of a monoclonal antibody that neutralized activated protein C (aPC) activity in this assay. **b**, Determination of aPC generation during the clotting of human plasma. Inset shows the amount of aPC bound to antibody-coated wells from samples withdrawn from clotting reactions run in the presence or absence of the thrombin inhibitor hirudin at the times indicated. The histogram shows the calculated amount of aPC generated in a 150- μ l clotting reaction using 60 nM FLIN-Q3 or wt HPC zymogens. **c**, Dose-response curve for the clot-activated activity of zymogen FLIN-Q3 compared to zymogen wt HPC. **d**, Determination of relative half-lives of activated wt HPC (aPC), activated FLIN-Q3 (aFLIN-Q3), and of FLIN-Q3 zymogen determined as activity generated in clotting plasma after incubation for various times before clot initiation.

METHODS. **a**, Wild-type HPC and FLIN-Q3 were added to 50 μ l human plasma at a concentration of 20 nM with 50 μ l Helena standard APTT (activated partial thromboplastin time) reagent and incubated for 5 min at 37 $^{\circ}$ C. Clotting of plasma was initiated by addition of 50 μ l 25 mM CaCl_2 to a final concentration of 8 mM, and clotting times measured. In concurrent experiments, monoclonal antibody HPC-1, which neutralizes aPC activity, was added to control plasma and to plasma containing either the zymogen wt HPC or FLIN-Q3. **b**, Amount of aPC generated during the clotting reaction was determined by the method described in ref. 28 in protein C-deficient plasma. Briefly, 96-well plates were coated with the HPC-3 monoclonal antibody against the light chain of protein C, blocked with a 5% solution of dry milk in phosphate-buffered saline (PBS), 0.1% sodium azide, and washed. Clotting reactions containing 60 nM FLIN-Q3 or wild-type zymogen with or without 10 units per ml hirudin were initiated by addition of CaCl_2 to a final concentration of 8 mM; samples were removed at various times, diluted in PBS containing 0.5% milk, 0.05% Tween 20 and 50 mM benzamidine, and incubated in the antibody-coated wells for 2 h at room temperature. Plates were washed extensively to remove benzamidine and the aPC activity was determined at 37 $^{\circ}$ C in a Thermomax kinetic plate reader using the chromogenic substrate S-2366 in 20 mM Tris, pH 7.4, 150 mM NaCl. Standard curves, using purified aPC diluted into plasma, gave correlation coefficients of 0.991 to 0.999, and were used to determine the amount of aPC bound to the well. **c**, Level of anticoagulant activity in clotting plasma was determined as a function of the concentrations of wt HPC and FLIN-Q3 zymogens, and the data were



expressed as prolongation of clotting time. Basal clotting times in the assay were from 27 to 33 s. **d**, Inhibition of each HPC derivative in plasma was determined by incubating normal human plasma (citrate) with 100 nM aPC, activated FLIN-Q3 (aFLIN-Q3), or zymogen FLIN-Q3 at 37 $^{\circ}$ C. Plasma concentration was 90% (v/v) in the final reaction in buffer containing 3 mM CaCl_2 , 150 mM NaCl, 20 mM Tris, pH 7.4, 1 mg ml⁻¹ BSA. Aliquots were removed at selected times, and, for aPC and aFLIN-Q3, activity was measured as amidolytic activity using S-2366 at a final concentration of 1 mM, or as the activated partial thromboplastin time³¹. For zymogen FLIN-Q3, samples were removed at various times and the aPC activity that could be generated in clotting plasma was determined. The $t_{1/2}$ values for aPC and aFLIN-Q3 were estimated from decay curves generated using Enzfitter Software (Elsevier Biosoft, UK).

to an increased affinity (decreased K_m) for thrombin and was independent of thrombomodulin (data not shown).

To test the concept that zymogen FLIN-Q3 might have anticoagulant activity at the site of a growing thrombus because of an enhanced sensitivity to cleavage, we investigated the amount of activated protein C generated from zymogen as a human clot was forming *in vitro*. For these experiments, we generated clots in human plasma in the presence and absence of wild-type protein C and FLIN-Q3 zymogens. As shown in Fig. 3a, the wild type had no anticoagulant effect as there was no increase in the normal clotting time of ~30 s observed in the control. But the zymogen FLIN-Q3 derivative prolonged the clotting time significantly. This effect was protein C-specific as antibodies raised against human protein C (either in sheep anti-HPC (American Diagnostic) or as monoclonal antibodies HPC-1 and HPC-3; ref. 27), effectively neutralized the anticoagulant activity generated from the zymogen. To ensure that this effect was due to protein C activation, we determined the level of activated protein C generated in the clotting plasma using an immunocapture assay²⁸. After initiation of the clotting reaction, a substantial amount of activated protein C was generated from FLIN-Q3 but not from human wild-type protein C (Fig. 3b). The generation of activated protein C from FLIN-Q3 could be completely blocked by the thrombin-specific inhibitor hirudin. As shown in the inset, the amount of activated protein C increased linearly with time, therefore the increase in clotting time appears to be due to the thrombin-induced generation of active FLIN-Q3. The ability of the zymogen derivative to prolong

the clotting time in this assay is striking, as enough activated FLIN-Q3 was generated almost simultaneously with the generation of thrombin. We estimated that from 5–10% of the zymogen became activated during the clotting reaction. As shown in Fig. 3c, the clot-activated activity of FLIN-Q3 was dose-dependent and linear with increasing concentration. The dose of zymogen FLIN-Q3 required to obtain a doubling of the clotting time was from 10–20 nM, compared to 1 to 2 nM with either activated protein C or activated FLIN-Q3. This difference corresponds to the 5–10% FLIN-Q3 activated during the reaction. Our results indicate that FLIN-Q3 has the potential to be a clot-targeted anticoagulant.

The activated form of protein C is rapidly inhibited in plasma by natural inhibitors, primarily the heparin-stimulated protein C inhibitor and α_1 -antitrypsin (see ref. 29). As shown in Fig. 3d, the activated form of FLIN-Q3 and the wild-type activated protein C had identical kinetics of inhibition, both had $t_{1/2}$ values of ~25 min in human plasma. The wild-type zymogen has a substantially longer $t_{1/2}$ than the activated protein C, partly because of its resistance to the action of natural inhibitors. Thus, we might expect zymogen FLIN-Q3 to have a similarly long half-life. As also shown in Fig. 3d, FLIN-Q3 retains the ability to be activated in clotting plasma even after relatively long incubation times before clot initiation.

We have demonstrated the ability to modify a natural substrate to generate a pro-enzyme that can be efficiently activated by the major procoagulant enzyme thrombin, resulting in an anticoagulant activity. If such an agent can be activated at the site

of arterial thrombin generation, then activation would not be confined to the microvasculature where thrombomodulin is present, which might be a clinical advantage. Also, lower doses of the thrombin-sensitive zymogen might be possible because of its longer half-life in plasma. The administration of an inactive, biologically stable agent that can be activated at the site of thrombin generation represents a novel and potentially safer approach to the treatment and prevention of thrombotic disease, although only clinical evaluation will establish the effects of mutations in protein C on immunogenicity, and on catabolism and clearance. □

Received 29 May; accepted 30 September 1992.

1. Esmon, N. L., Owen, W. G. & Esmon, C. T. *J. Biol. Chem.* **257**, 859–864 (1982).
2. Esmon, N. L., *Prog. Hemost. Thromb.* **9**, 29–55 (1989).
3. Esmon, C. T. *J. Biol. Chem.* **264**, 4743–4746 (1989).
4. Esmon, C. T. *Science* **235**, 1348–1352 (1987).
5. Gardner, J. E. & Griffin, J. H. *Human Protein C and Thromboembolic Disease* 265–278 (Grune and Stratton, New York, 1983).
6. Clouse, L. H. & Comp, P. C. *New Engl. J. Med.* **314**, 1298–1304 (1986).
7. Comp, P. C. & Esmon, C. T. *J. Clin. Invest.* **68**, 1221–1228 (1981).
8. Zolton, R. P. & Seegers, W. H. *Thromb. Res.* **3**, 23–27 (1973).
9. Romisch, J., Diehl, K.-H., Reiner, G. & Paques, E.-P. *Fibrinolysis* **5**, 191–196 (1991).
10. Gruber, A., Harker, L. A., Hanson, S. R., Kelly, A. B. & Griffin, J. H. *Circulation* **84**, 2454–2462 (1991).
11. Snow, T. R., Deal, M. T., Dickey, D. T. & Esmon, C. T. *Circulation* **84**, 293–299 (1991).
12. Esmon, C. T., Taylor, F. B. & Snow, T. R. *Thromb. Haemost.* **66**, 160–165 (1991).
13. Taylor, F. D. *et al. J. Clin. Invest.* **79**, 918–925 (1987).
14. Gruber, A., Griffin, J. H., Harker, L. A. & Hanson, S. R. *Blood* **73**, 639–642 (1989).
15. Dreyfus, M. *et al. New Engl. J. Med.* **325**, 1565–1568 (1991).
16. Okajima, K. *et al. Am. J. Hemat.* **33**, 277–278 (1990).
17. Kisiel, W. *J. Clin. Invest.* **64**, 761–769 (1979).
18. Emerick, S. C. *et al. in The Pharmacology and Toxicology of Proteins* 351–367 (Liss, New York, 1987).
19. Chang, J.-Y. *Eur. J. Biochem.* **151**, 271–224 (1985).
20. Feng, F. N., Meinwald, C., Vasquez, M. & Scheraga, H. A. *Biochemistry* **28**, 3094–3105 (1989).
21. LeBonniec, B. F. & Esmon, C. T. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7371–7375 (1991).
22. LeBonniec, B. F., MacGillivray, R. T. A. & Esmon, C. T. *J. Biol. Chem.* **266**, 13796–13803 (1991).
23. Grinnell, B. W. *et al. in Protein C and Related Anticoagulants* (eds Bruley, D. & Drohan, W.) 13–46 (Gulf, Houston, 1990).
24. Ehrlich, H. J. *et al. EMBO J.* **9**, 2367–2373 (1990).
25. Grinnell, B. W., Walls, J. D. & Gerlitz, B. *J. Biol. Chem.* **266**, 9778–9785 (1991).
26. Ye, J., Esmon, N. L., Esmon, C. T. & Johnson, A. E. *J. Biol. Chem.* **266**, 23016–23021 (1991).
27. Ehrlich, H. J., Jaskunas, S. R., Grinnell, B. W., Yan, S. B. & Bang, N. U. *J. Biol. Chem.* **264**, 14298–14304 (1989).
28. Gruber, A. & Griffin, J. H. *Blood* **79**, 2340–2348 (1992).
29. Heeb, M. J., Bischoff, R., Courtney, M. & Griffin, J. H. *J. Biol. Chem.* **265**, 2365–2369 (1990).
30. Parkinson, J. F. *et al. J. Biol. Chem.* **265**, 12602–12610 (1990).
31. Grinnell, B. W., Berg, D. T., Walls, J. & Yan, S. B. *Bio/Technology* **5**, 1189–1192 (1987).
32. Beckmann, R. J. *et al. Nucleic Acids Res.* **13**, 5233–5247 (1985).
33. Devereux, J., Haerberli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).

ACKNOWLEDGEMENTS. We thank B. Yan and N. Bang for discussion; J. Henry, C. Payne and Y. Roe for help with the HPC assay; B. Glover for assistance with oligonucleotide synthesis; and I. Jenkins for help with DNA sequencing.

Lifespan of human lymphocyte subsets defined by CD45 isoforms

Colin A. Michie*, Angela McLean†, Christopher Alcock‡ & Peter C. L. Beverley*

* Human Tumour Immunology Group, University College and Middlesex School of Medicine, 91 Riding House Street, London W1P 8BT, UK

† Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

‡ Department of Oncology and Radiotherapy, Churchill Hospital, Oxford OX3 7LJ, UK

THE lifespan of thymic-derived or T lymphocytes is of particular interest because of their central role in immunological memory. Is the recall of a vaccination or early infection, which may be demonstrated clinically up to 50 years after antigen exposure¹, retained by a long-lived cell, or by its progeny? Using the observation that T lymphocyte expression of isoforms of CD45 corresponds with their ability to respond to recall antigens, we have investigated the lifespan of both CD45R0 (the subset containing responders, or 'memory' cells) and CD45RA (the unresponsive, or 'naive' subset) lymphocytes in a group of patients after radiotherapy. Here we report rapid loss of unstable chromosomes from the CD45R0 but not the CD45RA pool. Immunological memory

therefore apparently resides in a population with a more rapid rate of division. Differing survival curves for the two subsets are best described by a model in which there is also reversion *in vivo* from the CD45R0 to the CD45RA phenotype. Expression of CD45R0 in T cells may therefore be reversible.

Estimates of the lifespan of human lymphocytes *in vivo* have been based on the persistence of unstable chromosomal lesions in peripheral circulating lymphocytes after radiotherapy. Dicentric chromosomal lesions with accompanying dissimilar chromosomal fragments result in cell death in mitosis, and their frequency therefore gives a measure of intermitotic time. Studies of a group of patients after radiotherapy demonstrated that 50% of cells with dicentric lesions disappeared from the periphery in a little less than 3 years (the lymphocyte half-life)². Using the correlations now established between phenotype and function, lymphocytes may be divided in order to investigate lifespan further. Both CD4 and CD8 human peripheral T lymphocytes can be separated into further subsets based on their expression of different isoforms of the CD45 (leucocyte common) antigen³. At birth, most T cells express isoforms of high relative molecular mass (M_r) which are detected with CD45RA monoclonal antibodies⁴. The number of cells expressing the lowest- M_r isoform detected by the CD45R0 antibody (UCHL1) increases thereafter, to reach the adult level of about 50% by the age of 10–20 years^{5,6}. Functional studies have shown that CD45R0 cells respond *in vitro* to recall antigens, whereas CD45RA positive cells do not⁷. This, together with the demonstration that CD45RA-positive cells express CD45R0 and lose CD45RA after mitogenic stimulation *in vitro*, has led to the suggestion that CD45RA are naive, and that CD45R0 are memory T lymphocytes^{8,9}. Although recent human and rat data call into question a straightforward association between expression of low- M_r isoforms and memory^{10–12}, the existence of CD45RA and CD45R0 subsets with very distinct properties prompted us to re-examine the question of their lifespan, and in particular to determine the survival of cells capable of responding to recall antigens.

We studied persistence of dicentric chromosomes in the

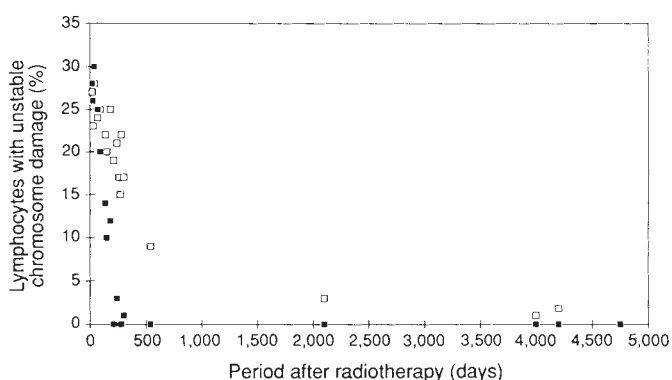


FIG. 1 Lymphocytes in each subset disappear at differing rates from the periphery. Nineteen consenting patients attending a radiotherapy follow-up clinic donated 20 ml venous blood for a full blood count (processed on a Technicon H-1 analyser) and investigation of their chromosomal morphology. All patients had received between 40 and 60 Gy radiation in fractionated doses over 2–4 weeks for the treatment of epithelial tumours; none had received chemotherapy or steroids. None received antibiotics or antiviral agents between radiotherapy and the time of sampling. Venous samples were separated immediately on Ficoll hypaque. After depletion of adherent cells, a T-lymphocyte preparation was made by removing cells expressing CD14, CD15, CD16 and CD20 antigens using monoclonal antibodies and magnetic beads⁷. CD45RA-positive (open squares) and CD45R0-positive T cells (filled squares) were purified from this preparation using magnetic beads. Only those samples containing less than 5% contaminants by FACS scan analysis were used to prepare chromosomes. From each venous sample, 100 sets of chromosomes were examined¹³, and counts made of the frequency of cells containing dicentric chromosomes with dissimilar chromosome fragments¹⁵.

CD45RA and CD45R0 subsets after therapeutic X-ray irradiation in a group of 19 cancer patients. At various times after irradiation, CD3+ CD45RA and CD45R0 lymphocytes were separated from peripheral blood samples and stimulated with phytohaemagglutinin. After mitotic arrest at 48 h with colchicine, the percentage of cells with dicentric chromosomes was determined (Fig. 1)^{13,14}. Two conclusions were drawn. First, the rate of disappearance of dicentric lesions from the two populations is very different. Thus by a year the dicentric lesions have been lost from the CD45R0 subset, whereas they persist to some 10 years in the CD45RA subset. A shorter half-life of lymphocytes

expressing the lower- M_r isoforms has also been observed in sheep¹⁵, whereas cells with a 'naïve' phenotype persist for long periods in mice after adoptive transfer¹⁶. The mean lymphocyte half-life calculated from our data, ignoring the division into subsets, is 729 days (95% confidence interval 584–971 days), approximating the published data¹. Second, steady loss, albeit at differing rates from the two populations, did not account for the observed pattern. This would produce exponential curves, which do not fit the data, particularly in the first 3 months after radiotherapy. A possible resolution of this difficulty is suggested by recent observations in rats. These show that T-cell subsets expressing either higher- M_r (CD45RC+) or lower- M_r (CD45RC-) isoforms of CD45 are capable of interconversion¹². Thus human lymphocytes may be capable of similar switching, from lower to higher- M_r isoforms of CD45. This led us to test a model in which cells of both subsets died, but in which cells were also allowed to revert from the CD45R0 to CD45RA phenotype without cell division. We assume also that cells do not convert from CD45RA to CD45R0 without cell division. Such a model generates survival curves fitting the observed data more closely (Fig. 2b). We suggest therefore that the rates of loss of dicentric chromosomes from the two subsets support the idea that reversion from the CD45R0 to CD45RA phenotype occurs *in vivo*.

How does a model in which CD45R0 memory cells appear to have a shorter lifespan accord with the long persistence of immunological memory? We have already proposed that memory is maintained by long-lived clones rather than individual cells with a long lifespan, and mechanisms capable of driving continued division in these clones have been suggested^{18–20}. Therefore a more puzzling question is that of the history of cells which in these patients revert from CD45R0 to RA phenotype. Are these cells that were originally recruited to the CD45R0 pool by antigen, or do some enter this pool during non-antigen-driven cell division in order to maintain the size of the peripheral lymphocyte pool, an idea supported by the rat experiments? Irrespective of what maintains the size of the total peripheral T-cell pool, mechanisms that maintain a T-lymphocyte clone in cell division, and concomitantly expressing CD45R0, are likely to be the most important factors determining the frequency of T lymphocytes responsive to a given recall antigen. The lifespan of the individual lymphocyte is probably not significant, but the persistence of CD45R0-positive progeny with the appropriate repertoire will determine the length of an immunological memory. Our data show that just as there is functional heterogeneity in the periphery, there is heterogeneity of intermitotic time (lifespan). Those lymphocytes responsible for recall responses have a more rapid turnover, implying that long-lived memory is not maintained by long-lived T-lymphocytes. □

Received 17 July; accepted 28 September 1992.

1. Couch, R. B. & Kasel, J. A. *Rev. Microbiol.* **37**, 529–549 (1983).
2. Buckton, K. E., Court Brown, W. M. & Smith, P. G. *Nature* **214**, 470–474 (1967).
3. Beverley, P. C. L. *Semin. Immunol.* **4**, 35–41 (1992).
4. Erkeller-Yuksel, F. et al. *J. Pediatr.* **120**, 216–221 (1992).
5. Hayward, A. R., Lee, J. & Beverley, P. C. *Eur. J. Immunol.* **19**, 771–773 (1989).
6. Terry, L. A., Pickford, A. & Beverley, P. C. L. in *Leucocyte Typing III* (eds McMichael, A. J. et al.) 225–227 (Oxford Univ. Press, Oxford, 1987).
7. Merckenschlager, M., Terry, L., Edwards, R. & Beverley, P. C. L. *Eur. J. Immunol.* **18**, 1653–1661 (1988).
8. Akbar, A. N., Terry, L., Timms, A., Beverley, P. C. & Janossy, G. *J. Immunol.* **140**, 2171–2178 (1988).
9. Merckenschlager, M. & Beverley, P. C. L. *Int. Immunol.* **1**, 450–459 (1989).
10. Warren, H. S. & Skipsey, L. *J. Immunol.* **74**, 78–85 (1991).
11. Rothstein, D. M., Yamada, A., Schlossman, S. F. & Morimoto, C. *J. Immunol.* **146**, 1175–1183 (1991).
12. Sparshott, S. M., Bell, E. B. & Sarawar, S. R. *Eur. J. Immunol.* **21**, 993–1000 (1991).
13. Hamerton, J. L. *Human Cytogenetics* (Academic, New York, 1971).
14. Buckton, K. E. & Pike, M. C. *Int. J. Rad. Biol.* **8**, 439–444 (1964).
15. Mackay, C. R., Marston, W. L. & Dudley, L. *J. exp. Med.* **171**, 801–817 (1990).
16. Sprent, J., Schaefer, M., Hurd, M., Surh, C. D. & Ron, Y. *J. exp. Med.* **174**, 717–728 (1991).
17. Raiston, M. L. & Jennerich, R. I. *Technometrics* **20**, 7–14 (1978).
18. Beverley, P. C. L. *Immun. Today* **11**, 203–205 (1990).
19. Melchers, F. in *Idiotypic Networks in Biology and Medicine* (eds Osterhaus, A. D. M. E. & Uytendaele, F. G. C. M.) 3–12 (Elsevier, Amsterdam, 1990).
20. Gray, D. & Matzinger, P. *J. exp. Med.* **174**, 969–974 (1991).

ACKNOWLEDGEMENTS. We thank G. Janossy for antibodies. This work was supported by the Imperial Cancer Research Fund. A. M. is supported by a Royal Society Fellowship. The study was registered by the Oxford Central Ethical committee.

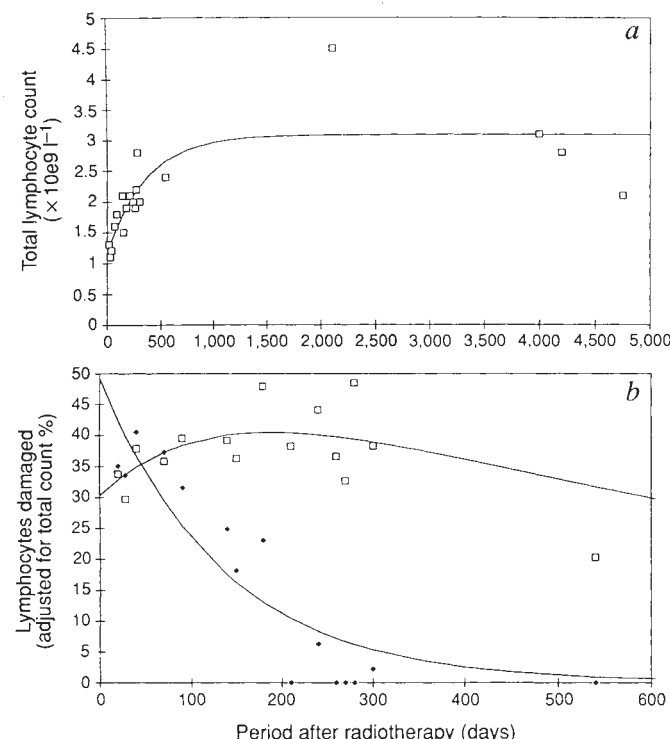


FIG. 2 Lymphocytes of CD45R0 phenotype can revert to CD45RA phenotype. *a*, Data recording the per cent of lymphocytes damaged over time does not take into account the fact that in the period immediately after radiotherapy, the peripheral lymphocyte count in the majority of patients falls. Shown are total lymphocyte counts of the patient group and the best fit curve $y = p_1 - p_2 \exp(-p_3 t)$. Best estimates (± 2 s.e.) for these variables are $p_1 = 3.09$ (± 0.49), $p_2 = 1.95$ (± 0.75), $p_3 = 0.0028$ (± 0.0024). We used the predicted total lymphocyte count to calculate the absolute numbers of lymphocytes in each subset in *b*. The numbers of lymphocytes of CD45R0-positive phenotype with unstable chromosomal damage (filled diamonds) follows an exponential decay with time, whereas those with CD45RA-positive phenotype and unstable damage follow a differing decay pattern (open squares) with an increase in cell numbers, followed by a decrease. Given that all cells with unstable chromosomal damage die on cell division, the increase can only be accounted for by reversion of CD45R0-positive lymphocytes to CD45RA-positive phenotype. This possibility was formally tested using nonlinear parameter estimation. Curves predicting the numbers of damaged lymphocytes were constructed so as to allow for three possibilities: CD45RA cells die, CD45R0 cells die and CD45R0 cells change into CD45RA cells. These curves are the solution to the following pair of differential equations which describe the populations of naïve (A) and memory (O) lymphocytes: $dA/dt = cO - \mu_A A$ and $dO/dt = -(c + \mu_O)O$. The five parameters required for these curves were derived from the data: the initial numbers of damaged cells in each subset, the death rate for each subset, and the conversion rate from one subset to the other¹⁷. If there is conversion from memory to naïve phenotype, the conversion rate will be significantly greater than zero. The best estimates (± 2 s.e.) for the death rates from these data were 3.8 (± 0.3) per 1,000 CD45R0 cells per day; 1.1 (± 0.8) per 1,000 CD45RA cells per day; the conversion rate was 3.6 (± 2.8) per 1,000 CD45R0 cells per day. As the conversion rate is significantly different from zero, we must reject the possibility that both populations decay exponentially, and conclude that CD45R0 phenotype cells do revert to CD45RA phenotype *in vivo*. The lines on the graph show predicted population sizes using the best estimates of the parameter values.

Signalling through the MHC class II cytoplasmic domain is required for antigen presentation and induces B7 expression

N. Nabavi*, G. J. Freeman†, A. Gault*, D. Godfrey*, L. M. Nadler† & L. H. Glimcher‡§

* Department of Immunopharmacology, Roche Research Center, Nutley, New Jersey 07110, USA

† Division of Tumor Immunology, Dana Farber Cancer Institute, and

‡ Harvard University School of Public Health and Harvard Medical School, 665 Huntington Avenue, Boston, Massachusetts 02115, USA

§ To whom correspondence should be addressed

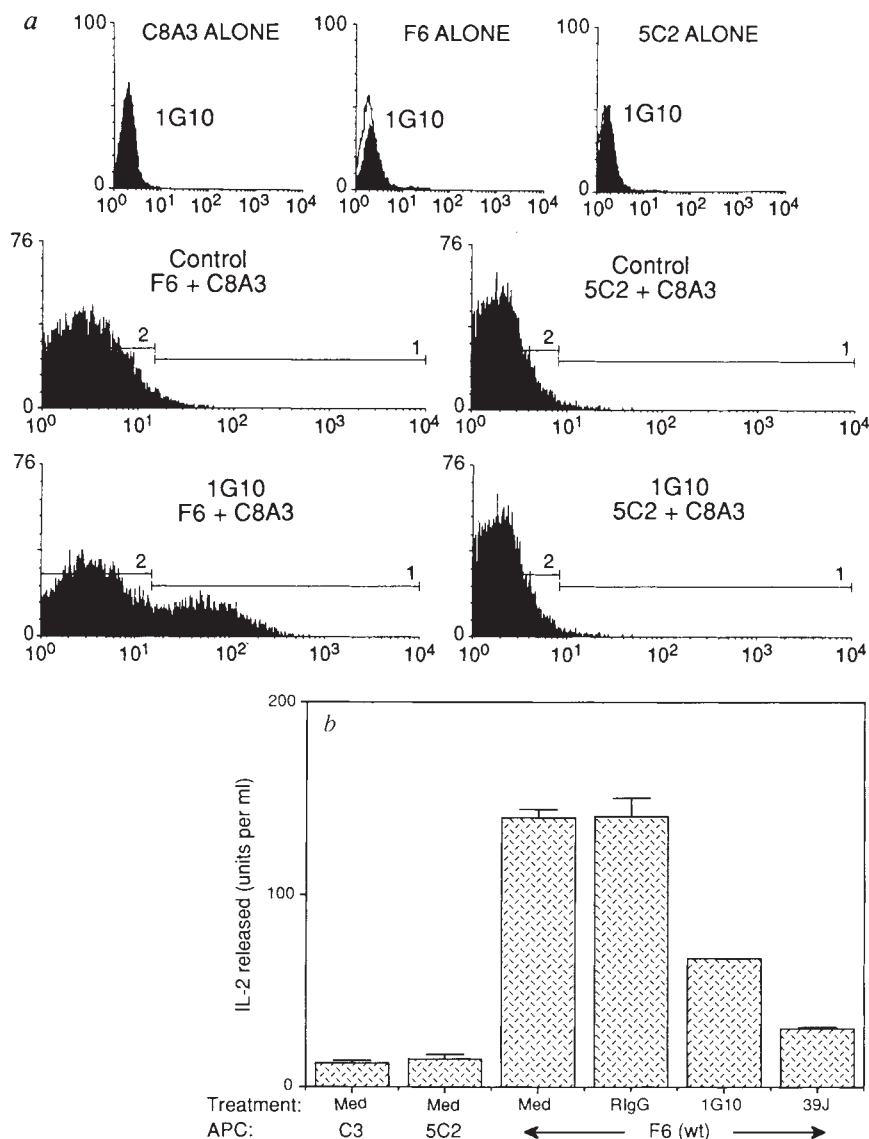
CLASS II major histocompatibility complex (MHC) molecules function as antigen-presenting elements as well as signal transducers on B lymphocytes. We previously reported that a B lymphoma cell transfectant, 5C2, expressing genetically engineered I-A^k molecules with truncated cytoplasmic domains was severely impaired in both antigen presentation and in anti-Ia-induced intracytoplasmic signalling¹⁻⁵. These two functions could be restored by preculturing 5C2 cells with cyclic AMP analogues^{6,7}.

Here we demonstrate that impaired signal transduction by truncated class II molecules results in a deficiency in induction of the newly defined B-cell accessory molecule B7 (ref. 8), which can be reversed by restoration of B7 expression. These data imply that contact of the T-cell antigen receptor with MHC/antigen ligand results in signal transmission through the class II cytoplasmic domain. This signal, which can be mimicked by dibutyryl cAMP, induces expression of B7, resulting in effective antigen presentation. The fact that crosslinking of surface class II MHC also induces B7 expression on normal resting human B cells⁹ supports this contention.

When assayed for antigen-presentation ability, 5C2 cells stimulated the CD28-expressing autoreactive T hybrid, C8A3, to secrete less than 1% of the interleukin-2 secreted in response to F6 cells (Fig. 1b; refs 1 and 6). To determine whether the recently defined CD28/B7 accessory pathway was involved in C8A3 activation, F6 or 5C2 cells were co-cultured with C8A3 cells and subsequently examined by flow cytometry (FACS) for expression of B7 antigen using a monoclonal antibody, 1G10, recently developed against murine B7 (N.N. and D.G., unpublished results). B7 molecules were not present on the surface of unstimulated F6 cells but were induced upon contact with T hybrid cells (Fig. 1a). Neither C8A3, F6 cells or 5C2 cultured alone express surface protein (Fig. 1a, upper panels) or B7 messenger RNA (Fig. 4b): B7 antigen and B7 mRNA (not shown) is induced only upon co-culture of F6 antigen-presenting

FIG. 1 a, Coculture of C8A3 with F6 but not with 5C2 induces B7 surface expression. The control staining is superimposed on the 1G10 staining for cells cultured alone (upper panel), but is shown separately for the co-cultured cells (lower panels). **b**, Antigen presentation by F6 cells is blocked by a monoclonal antibody against murine B7.

METHODS. The 5C2 cell was derived by transfecting A_β^k and A_α^k chain genes lacking 10 and 12 carboxy-terminal amino acids of the cytoplasmic domain, respectively, into the Ia-negative B lymphoma M12.C3 (C3). The resulting cell line expressed levels of I-A^k comparable to its wild-type I-A^k-expressing partner, F6 (ref. 1). C8A3 T hybrid cells were co-cultured at a 1:1 ratio with wild-type (wt) F6 or truncated 5C2 cells at 10⁶ per ml for 40 h and then stained with monoclonal antibody (mAb) 1G10 coupled with fluorescein isothiocyanate (FITC) or with a control FITC-conjugated rat mAb before being analysed by flow cytometry. Inhibition of antigen presentation by F6 cells was observed by co-culturing 1 × 10⁵ C8A3 autoreactive I-A^k-restricted T-cell hybrids with 2 × 10³ F6 cells in a total volume of 200 μl. After 16 to 18 h, serial dilutions of the supernatants of the cultures were assayed for the presence of interleukin-2 (IL-2) by measuring incorporation of tritiated thymidine into the IL-2-dependent T-cell line CTLL (ref. 1); units per ml of IL-2 were determined by comparison with a standard curve using recombinant IL-2. IL-2 units are expressed as a mean value ± s.e.m. of duplicated samples from one representative experiment out of three. The per cent inhibition of antigen presentation by 1G10 was determined by comparison to the F6 response. A final concentration of 20 μg ml⁻¹ of 1G10, an isotype-matched control rat mAb (RlgG), or 4.8 μg ml⁻¹ of 39J (anti I-A^k mAb) was added at the onset of culture.



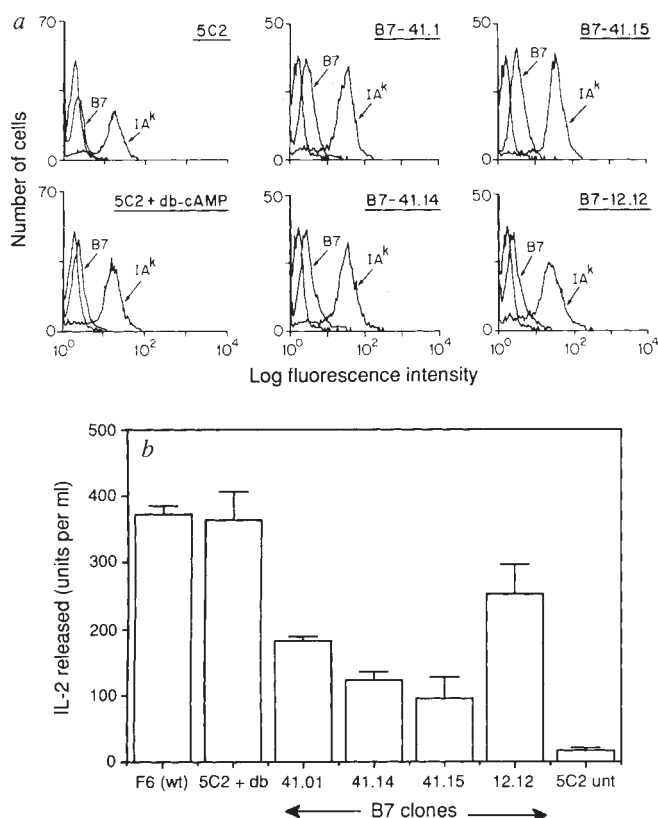


FIG. 2 The expression of exogenous B7 in 5C2 cells restores antigen presentation. *a*, Untreated 5C2 controls, dibutyl (db) cAMP-induced 5C2 cells, and four different 5C2-B7 transfected clones were stained with both biotinylated-1G10 coupled with phycoerythrin-avidin, and FITC-labelled 11-5.2, anti-I-A^k (Pharmingen) before being analysed by flow cytometry. *b*, Ability of the four 5C2-B7 clones to present antigen to the autoreactive I-A^k-restricted T hybrid, C8A3. Antigen presentation by the same number of db-cAMP-induced (db) 5C2 cells is shown for comparison. 5C2 unt, Untreated 5C2 cells. Assay procedure is the same as for Fig. 1*b*, except that 1×10^4 APCs were used.

METHODS. The entire coding region of the murine B7 cDNA¹⁰ was subcloned into the pRc/CMV (cytomegalovirus) expression vector (Invitrogen), colinear with the neomycin-resistance gene. Recombinant plasmid DNA (25 μ g) was linearized with the restriction enzyme *PvuI* and transfected into 5C2 cells by electroporation. 5C2 cells transfected with the expression vector alone without the B7 cDNA insert served as controls. Transfectants were selected in the presence of 350 μ g ml⁻¹ G418, subcloned twice and were screened for surface B7 expression by flow cytometry using 1G10 mAb.

cells (APC) with the T hybrid (Fig. 1*a*, left-hand lower panels). That the signal for inducing B7 expression is transmitted through the cytoplasmic domain of the class II molecule is demonstrated by the failure of C8A3 to induce B7 surface expression (Fig. 1*a*, right-hand lower panels) or B7 mRNA (not shown) in the 5C2 transfectant. Double-colour FACS with an anti-CD3 monoclonal antibody identified the negative population of cells as T cells (not shown). This experiment demonstrates that interaction of class II with a T-cell-surface molecule, presumably the T-cell receptor, leads to signal transmission through the class II cytoplasmic domain, resulting in B7 gene transcription. To test directly whether B7 expression was functionally involved in antigen presentation, we used the 1G10 antibody in blocking experiments. Antigen presentation by F6 to C8A3 was blocked by the presence of 20 μ g ml⁻¹ of 1G10 (Fig. 1*b*). This blockade is partial, as may be expected if other accessory pathways exist that are important (Fig. 4), or if the 1G10 antibody is only a moderately efficient blocking antibody (Fig. 3*a*). Indeed, in preliminary experiments, other monoclonal antibodies generated against dibutyl cAMP-treated 5C2 cells, but not specific for B7 or class II, partially block antigen presentation, so providing evidence for additional accessory pathways (N.N., unpublished data).

We next tested whether the antigen-presentation defect could be corrected by transfecting 5C2 cells with a murine B7 complementary DNA¹⁰ driven by a heterologous promoter. Figure 2*a* shows FACS profiles of four different 5C2-B7 transfected clones positive for 1G10 binding. Interestingly, the 5C2-B7 transfectants did not express high levels of surface B7. Nevertheless, expression of B7 in 5C2 cells restored antigen presentation to C8A3 up to 67% compared with wild-type F6 (Fig. 2*b*). Antigen presentation by both wild-type F6 (Fig. 1*b*) and 5C2-B7 transfectants (Fig. 3*a*) was blocked in the presence of 20 μ g ml⁻¹ of 1G10. The per cent inhibition by the 1G10 antibody ranged from ~55 to 90% with a fixed number of APC in several experiments, suggesting that this antibody was only moderately

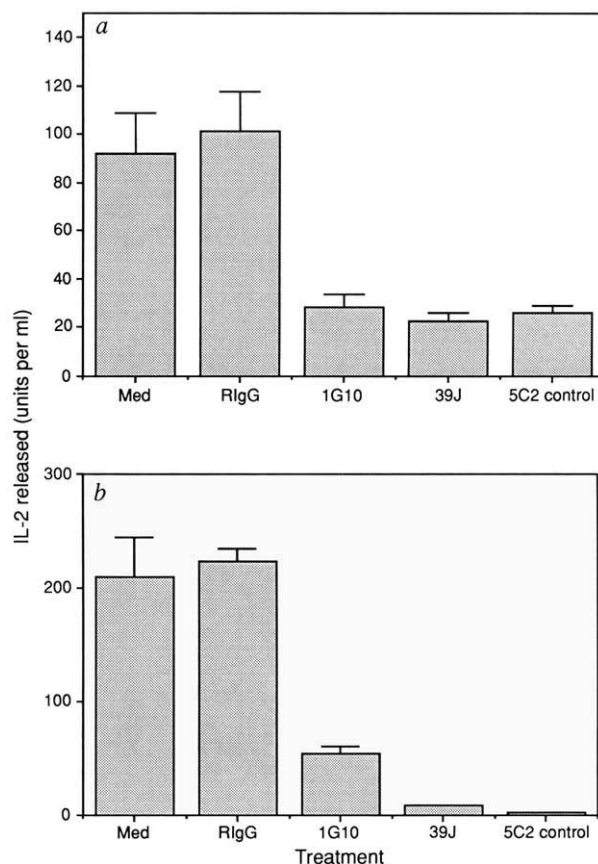


FIG. 3 *a*, mAb 1G10 blocks antigen presentation by 5C2-B7 transfectants. Antigen presentation was assayed as outlined in Fig. 1 legend, except 5×10^3 5C2-B7 (12.12) cells were used. *b*, Antibody 1G10 blocks antigen presentation activity of db-cAMP-induced 5C2 cells.

METHODS. 5C2 B cells were precultured in the presence of 300 μ g ml⁻¹ db-cAMP (Boehringer Mannheim) overnight. Cells were washed three times and co-cultured at 5×10^3 cells with 1×10^5 cells of the I-A^k-restricted autoreactive T cell hybrid, C8A3, in a total volume of 200 μ l. Experimental procedures and final concentrations of antibodies were as described in Fig. 1*b*.

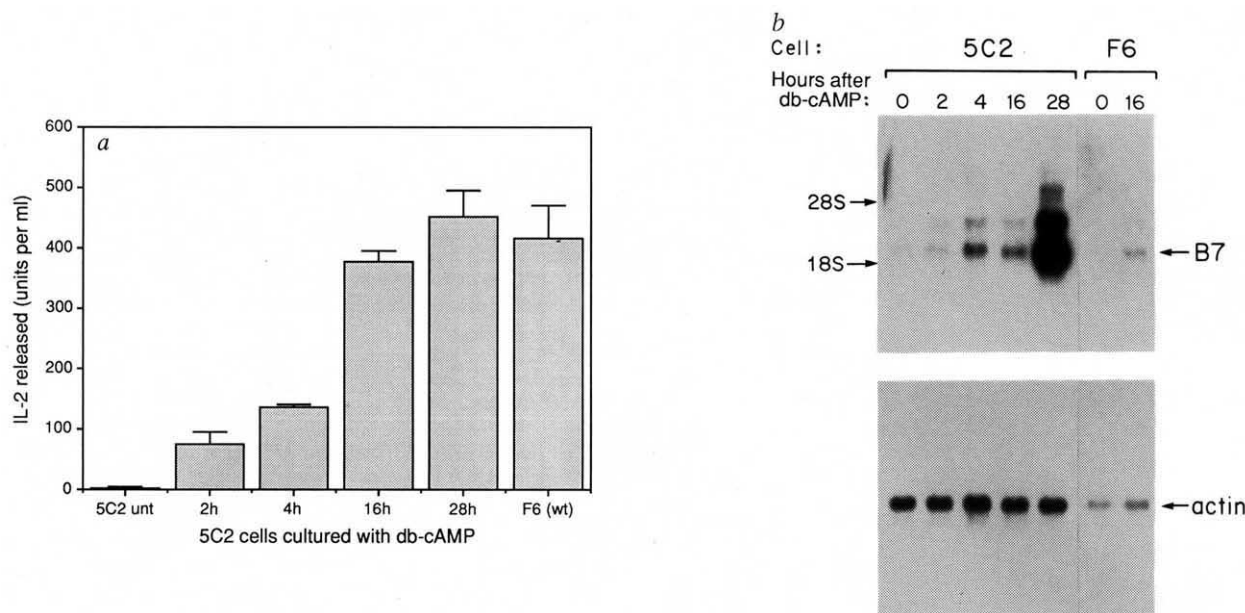


FIG. 4 Time course of action of db-cAMP on the 5C2 B cell line. *a*, Restoration of antigen presentation in 5C2 cells by db-cAMP. *b*, Kinetics of B7 mRNA induction by db-cAMP.

METHODS. 5C2 cells were precultured with 300 $\mu\text{g ml}^{-1}$ db-cAMP for the time indicated. Experimental procedures were as described in Fig. 3b, except that 1×10^5 5C2 cells were used. IL-2 activity was measured and presented as in Fig. 1b. The efficiency of antigen presentation by 5C2 cells was compared to the wild-type (wt) I-A^k-expressing M12.C3 transfectant cell line F6. These data represent the mean percentage of wt F6 activity $\pm$ s.d.

efficient at blocking. No correlation between surface levels of B7 and antigen presentation function in the various 5C2-B7 transfectants was obvious, suggesting that a small number of B7 molecules is sufficient to provide a co-stimulatory signal to T cells, consistent with other reports that levels of B7 expression on B cells are very low⁹.

Dibutyl cAMP treatment of 5C2 cells induces surface B7 expression and antigen presentation like the 5C2-B7 transfectants (Fig. 2a and b). To explore further the relationship of signal transduction through the cytoplasmic domain of class II MHC to a cAMP-driven pathway, 5C2 cells were precultured with dibutyl cAMP and their antigen-presentation capacity and the induction of B7 was tested. Dibutyl cAMP-treated 5C2 cells acquired wild-type antigen-presentation capacity (ref. 6, and Fig. 4a). The earliest effects of dibutyl cAMP on antigen-presentation function of 5C2 cells are detectable by two hours (Fig. 4a). Antigen presentation was partly (but not completely¹⁰) blocked by $\geq 20 \mu\text{g ml}^{-1}$ 1G10 (Fig. 3b). Northern analysis of RNA extracted from 5C2 cells cultured with dibutyl cAMP for varying lengths of time reveals that dibutyl cAMP induces transcripts for the murine B7 molecule by 2 h, which increased maximally by 28 h (Fig. 4b). The relative increase in

of duplicated samples from one out of five independent experiments. Total RNA was extracted from the same cell population used in the antigen presentation assay. Poly(A)⁺ RNA (1 μg) prepared from each sample was electrophoresed on an agarose-formaldehyde gel and blotted onto nitrocellulose (Schleicher and Schuell). Blots were hybridized to a 921-bp murine B7 coding region polymerase chain reaction product and subsequently with an actin cDNA probe, essentially as before¹⁰. Autoradiographs represent 24-h and 5.5-h exposures for B7 and actin, respectively. Migration of 18S and 28S ribosomal RNA markers is indicated.

levels of B7 mRNA correlated well with the extent of restoration of antigen presentation by these cells (Figs 4a and b).

In vitro studies indicate that the interaction of B7 (refs 11–13) on activated APC with its receptor CD28 (refs 8, 14 and 15) on T cells is required as a second signal for maximum cytokine production, proliferation of CD4⁺ T-helper cells and cell-mediated cytotoxicity^{14–19}. Certain T hybrids bearing high-affinity T-cell antigen receptors may not require co-stimulatory signals. For example, some T cells can be efficiently stimulated in the absence of B7 (ref. 1), CD4 or CD8 accessory molecules²⁰. Our data suggest that during cognate T-B-cell interaction, engagement of the T-cell receptor with class II MHC and antigen delivers a signal to the B cell. This signal transduction requires the cytoplasmic domain of the class II molecule and results in surface B7 expression. It is evident, therefore, that in addition to T-cell receptor interaction with MHC class II antigen complexes, a successful outcome of T-helper cell/B cell collaboration requires the participation of additional transmembrane receptor ligand pairs. Some of these are constitutively expressed, whereas others like B7 are probably induced by signals transmitted by class II MHC/T-cell receptor interactions. □

Received 21 April; accepted 16 September 1992.

- Nabavi, N. *et al.* *J. Immunol.* **142**, 1444–1447 (1989).
- Cambier, J. C. *et al.* *Nature* **327**, 629–632 (1987).
- Wade, W. F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 6297–6301 (1989).
- Pollok, K. E. *et al.* *J. Immunol.* **146**, 1633–1641 (1991).
- Chen, Z. Z., McGuire, J. C., Leach, K. L. & Cambier, J. C. *J. Immunol.* **138**, 2345–2352 (1987).
- St-Pierre, Y., Nabavi, N., Ghogawala, Z., Glimcher, L. H. & Watts, T. H. *J. Immunol.* **143**, 808–812 (1989).
- St-Pierre, Y. & Watts, T. H. *J. Immunol.* **147**, 2875–2882 (1991).
- June, C. H., Ledbetter, J. A., Linsley, P. S. & Thompson, C. B. *Immun. Today* **11**, 211–216 (1990).
- Koulova, L., Clark, E. A., Shu, G. & Dupont, B. *J. exp. Med.* **173**, 759–762 (1991).
- Freeman, G. J. *et al.* *J. exp. Med.* **174**, 625–631 (1991).
- Freedman, A. S., Freeman, G. J., Rhyhart, K. & Nadler, L. M. *Cell Immunol.* **137**, 429–437 (1991).
- Yokochi, T., Holly, R. D. & Clark, E. A. *J. Immunol.* **128**, 823–827 (1982).

- Freedman, A. S., Freeman, G., Horowitz, J. C., Daley, J. & Nadler, L. M. *J. Immunol.* **139**, 3260–3267 (1987).
- Linsley, P. S. *et al.* *J. exp. Med.* **173**, 721–730 (1991).
- Linsley, P. S. *et al.* *J. exp. Med.* **174**, 561–569 (1991).
- Linsley, P. S., Clark, E. A. & Ledbetter, J. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5031–5035 (1990).
- Gimmi, C. D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 6575–6579 (1991).
- Damle, N. K., Linsley, P. S. & Ledbetter, J. A. *Eur. J. Immunol.* **21**, 1277–1281 (1991).
- Reiser, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **89**, 271–275 (1992).
- Marrack, P. *et al.* *J. exp. Med.* **158**, 1077–1091 (1983).

ACKNOWLEDGEMENTS. We thank M. Grusby, T. Laufer, J. Markowitz, A. Singer, D. Singer and G. Powers for reviewing the manuscript, and J. Allison for testing CD28 expression on the C8A3 hybrid. Supported by USPHS (L.H.G. and L.M.N.) and the Leukemia Society Scholar Award (L.H.G.).

Rat liver protein linking chemical and immunological detoxification systems

Frank A. Blocki^{*†}, Patrick M. Schlievert[‡]
& Lawrence P. Wackett^{*†‡§}

^{*} Gray Freshwater Biological Institute and Department of Biochemistry, University of Minnesota, P.O. Box 100, Navarre, Minnesota 55392, USA

[†] Present address: Department of Biochemistry and Institute for Advanced Studies in Biological Process Technology, University of Minnesota, St Paul, Minnesota 55108, USA

[‡] Department of Microbiology, University of Minnesota, Box 196, Minneapolis, Minnesota 55455, USA

[§] To whom correspondence should be addressed

MAMMALS have separate enzymatic and cellularly mediated detoxification systems. Glutathione *S*-transferases (GSTs) protect against xenobiotic chemicals which continuously enter the body, largely through mucous membranes¹⁻⁵. These enzymes catalyse the conjugation of glutathione with a wide variety of electrophilic compounds rendering them non-toxic. Mammals also mount a cellular immunological response on entry of foreign cells, viruses or macromolecules into the body⁶⁻⁸. T lymphocytes mobilize at the site of foreign body entry and secrete protein messengers called lymphokines. Secondary to T lymphocytes, macrophages concentrate at the infection site and function in antigen processing and phagocytosis. *In vitro*, macrophage movement is arrested by one class of lymphokines known as macrophage migration inhibitory factors (MIFs). We report here the purification of milligram quantities of a unique multifunctional protein from rat liver which links enzymatic and immunological detoxification systems. This protein acts as both GST and MIF activity and matches the primary structure of a human MIF⁹ in 25 out of 26 amino-terminal amino acids. Primary structure comparisons revealed significant similarity between GSTs and MIF. The glutathione affinity chromatography purification described here yields a 100-fold increase in obtaining MIF⁹⁻¹¹ and will aid understanding of its precise biological function.

TABLE 1 Migration inhibition of rat nonimmune peritoneal exudate cells (PEC) by the 13K protein

Protein added to assay	Protein concentration ($\mu\text{g ml}^{-1}$)	Migration image weight (mg)	Inhibition (%)
TSST	0	74.6 $\pm$ 5.8 (4)	0
TSST	2	34.7 $\pm$ 4.2 (3)	54
TSST	10	13.4 $\pm$ 0.8 (3)	82
13K protein	0	90.0 $\pm$ 8.1 (4)	0
13K protein	2	53.6 $\pm$ 18.3 (3)	55
13K protein	7	33.7 $\pm$ 2.1 (3)	72

PEC migrations (80,000 cells per 100 μl capillary) were traced from a projected image, cut out, and weighed. Toxic shock syndrome toxin (TSST) was used as a positive control¹⁵. The number of replicates are given in parentheses. Total protein quantifications were made with the Pierce BCA (bicinchoninic acid) protein assay²². Here, the 13K protein may function either as an inducer of MIF action (like TSST) or directly as MIF. Methods. Three days after peritoneal injection of 10 ml mineral oil into a 350 g Sprague-Dawley rat, PEC²³ were collected with a Penrose drain after intraperitoneal lavage with 50 ml RPMI medium (Gibco). This medium contains 25 mM HEPES buffer, 200 mM L-glutamine, 100 U penicillin, and 100 g streptomycin per ml supplemented with 2% fetal calf serum. Aspirates (45 ml) were centrifuged at 500g for 15 min, washed three more times in 10 ml RPMI before final suspension in 1.2 ml RPMI. PEC were drawn up into 100 μl heparinized capillaries, plugged with clay, spun and cut, and the cell-containing portions were fastened to the bottom of 24-well microtitre plates with vasoline as described by David¹². Wells were overlaid with 500 μl RPMI with or without TSST or 13K protein. The plates were incubated at 37 °C and the extent of macrophage migration was determined at 32 h.

In the course of experiments to identify the rat liver GST isozymes active in the detoxification of dichloromethane, a novel protein with a subunit M_r of 13K was purified to homogeneity (Figs 1 and 2). The procedure yielded 1 mg of the 13K protein per 15 g of soluble liver protein. The 13K protein catalysed the conjugation of glutathione with 1,2-epoxy-3-(*p*-nitrophenoxy) propane with a specific activity of 7 $\mu\text{mol min}^{-1}$ per mg of protein. The 13K protein is not significantly active with 1-chloro-2,4-dinitrobenzene, the substrate most often used in the assay of GSTs. This and the stringent conditions required to elute it from the CM-52 resin may explain why the 13K protein has not been previously described.

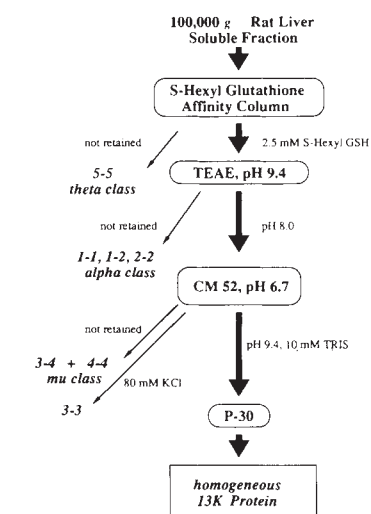
All previous reports of mammalian GSTs describe hetero- or homodimers with subunit M_r s of 24K to 29K¹⁻⁵. In this case, GST activity was unambiguously shown to be associated with the 13K protein as follows. (1) Gel filtration using P-30 resin yielded an enzymatically active uniform protein peak corresponding to an M_r of 13K and separated by 50 fractions from a minor higher M_r GST (Fig. 2). (2) SDS-polyacrylamide gel electrophoresis (PAGE) showed a significant band corresponding to an M_r of 13K with no discernible contaminating bands (Fig. 2 inset). (3) Dialysis using a 12K cutoff membrane resulted in loss of both protein and activity, concomitantly.

N-terminal sequencing of the 13K protein revealed that 25 out of 26 amino-acid residues directly matched those derived from sequencing a human MIF complementary DNA (Fig. 3a)⁹. This human MIF is reported to have a M_r of 12,650. The MIF biological activity of the 13K protein was determined in the standard capillary assay¹²⁻¹⁴. Macrophage migration was inhibited by 55 and 72% with 2 and 7 μg of the 13K protein, respectively (Table 1). This compares favourably with the activity induced by purified toxic shock syndrome toxin (TSST), which potentially stimulates T cells to secrete MIF¹⁵. Thus, the 13K GST has both structural and apparent functional characteristics of one of the MIF family of peptides.

Further comparison of the 13K protein with GSTs revealed additional primary structural similarity. The 13K protein's 26 N-terminal residues displayed a 35% direct sequence homology

FIG. 1 GST and 13K protein isolation flow chart. Isolation protocols were adapted from Boyer²⁴ and Meyer²⁵. Column procedures are ringed, isozymes are in italics, and the elution conditions for the 13K protein are traced out by the bold-faced arrows. Minor arrows show where previously described GSTs elute. Isozymes were identified by their substrate specificity and by N-terminal sequencing. GSH is glutathione.

METHODS. Sprague-Dawley rats were fed standard Purina rat chow, anaesthetized with ether; their livers were surgically removed and immediately frozen in liquid nitrogen. Livers were thawed in a 10 mM potassium phosphate buffer, pH 7 and the soluble fraction prepared as described by Meyer²⁵. Activity was determined with 0.5 mM 1,2-epoxy-3-(*p*-nitrophenoxy) propane (EPNP) under the conditions described by Fjellsted²⁶. In control incubations, no background reaction was detected at pH 6.5. The glutathione transferase-dependent transformation of dichloromethane to formaldehyde²⁷ was determined with a coupled assay using formaldehyde dehydrogenase²⁸. Dichloromethane dehalogenase activity of the 13K protein was proportional to the protein concentration and showed saturation effects with dichloromethane. The apparent K_m for dichloromethane and the V_{max} were 700 μM and 9 nmol min^{-1} per mg protein, respectively, at a fixed glutathione concentration of 1 mM. Similar experiments with EPNP were not done because of its low water solubility.



	1	5	10	15	20	25																				
Human MIF	Met	Pro	Met	Phe	Ile	Val	Asn	Thr	Asn	Val	Pro	Arg	Ala	Ser	Val	Pro	Asp	Gly	Phe	Leu	Ser	Glu	Leu	Thr	Gln	Gln
13K Protein	Met	Pro	Met	Phe	Ile	Val	Asn	Thr	Asn	Val	Pro	Arg	Ala	Ser	Val	Pro	Glu	Gly	Phe	Leu	Ser	Glu	Leu	Thr	Gln	Gln
Rat GST 3-3	Met	Pro	Met	Ile	Leu	Gly	Tyr	Trp	Asn	Val	Arg	Gly	Leu	Thr	His	Pro	Ile	Arg	Leu	Leu	Leu	Glu	Tyr	Thr	Asp	Ser

FIG. 3 Primary structural comparisons between the 13K protein (this study), human MIF⁹ and rat GST 3-3¹⁶. Direct matches are surrounded by boxes and conservative substitutions (defined in ref. 17) are underlined.

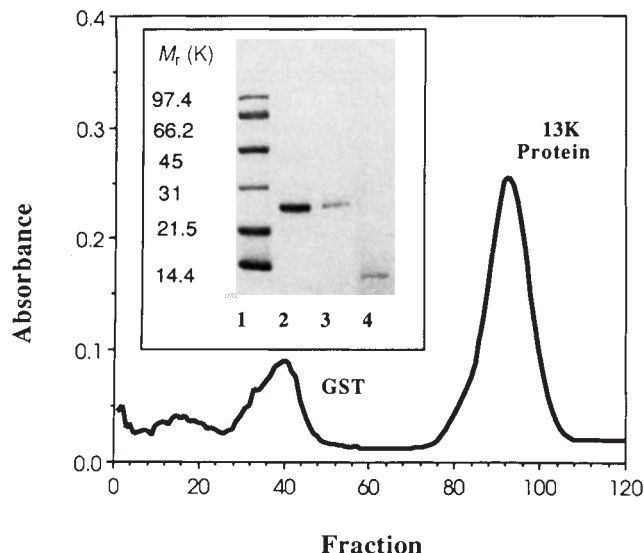


FIG. 2 P-30 gel exclusion chromatography and SDS-PAGE (inset). P-30 column fractions were monitored by absorbance at 280 nm. The elution volumes were: GST, 60 ml and 13K protein, 152 ml. Inset, 15% SDS-PAGE with lane 1, Bio-Rad low M_r markers; lane 2, CM 52 breakthrough, GSTs 3-4 and 4-4; lane 3, CM 52 gradient, GST 3-3; and lane 4, purified 13K protein. The M_r of the marker proteins are indicated on the left.

with the gene sequence-derived primary sequence of rat liver GST 3-3¹⁶ (Fig. 3). Of the remaining positions, 38% represented conservative substitutions¹⁷. Recent X-ray structure determinations of GSTs highlight the significance of N-terminal residues in the catalytic mechanism¹⁸⁻²⁰.

These data bring out new structure-function correlations relating to biological GSTs and MIF. The sequenced human MIF is 115 residues long⁹ and the rat 13K protein is of similar size. Several previous studies suggest that structures smaller than the known GST dimers might be sufficient to provide both glutathione and electrophilic cosubstrate binding sites and to catalyse nucleophilic attack by the glutathione thiolate. For example, each GST subunit is kinetically independent in catalysis²¹ and domain I of pig π GST (residues 1-74) contains the glutathione binding site¹⁸. The binding site for the electrophilic cosubstrate is less well defined at present. Future elucidation of the three-dimensional structure of the MIF family of proteins will contribute to the knowledge of the minimal structural requirements for GST catalysis and the biological mechanisms of MIF action. It is intriguing to speculate that GST activity of the 13K protein might be linked with its physiological function as an MIF. Further studies with active-site modifying reagents may help resolve this question. The convenient MIF purification protocol reported here should open the way to more detailed studies on the connections between enzymatic and cellularly mediated self-protection systems. □

- Jakoby, W. B. in *Glutathione S-transferases and Drug Resistance* (eds Hayes, J. D., Pickett, C. B. & Mantle, T. J.) 87-96 (Taylor and Francis, London, 1990).
- Armstrong, R. N. *Chem. Res. Toxicol.* **4**, 131-140 (1991).
- Goust, J. M. *Immun. Series* **28**, 169-189 (1986).
- Adams, D. O. & Hamilton, T. A. *Immun. Rev.* **97**, 5-27 (1987).
- Allison, A. C. *Immun. Rev.* **40**, 3-27 (1978).
- Weiser, W. Y. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7522-7526 (1989).
- Oki, S., Hirose, S. I., Higuchi, M. & Osawa, T. *Lymph. Cyt. Res.* **10**, 273-280 (1991).
- Hirose, S. H., Oki, S., Higuchi, M. & Osawa, T. *Microbial Immun.* **35**, 235-245 (1991).
- David, J. R. *Proc. natn. Acad. Sci. U.S.A.* **56**, 73-77 (1966).
- Bloom, B. R. & Bennett, B. *Science* **153**, 80-82 (1966).
- Rajapakse, D. A. & Glynn, L. E. *Nature* **226**, 857-858 (1970).
- Schlievert, P. M. *J. infect. Dis.* **147**, 391-398 (1983).
- Ding, G. J. F., Lu, A. Y. H. & Pickett, C. B. *J. biol. Chem.* **260**, 13268-13271 (1985).
- Lathrop, R. H., Webster, T. A. & Smith, T. F. *Ariadne Manual* (Odyssey Biosoft, 10 Lanark Road, Arlington, MA 02174, 1989).
- Reinemer, P. et al. *EMBO J.* **10**, 1997-2005 (1991).
- Liu, S. et al. *J. biol. Chem.* **267**, 4296-4299 (1992).
- Zhang, P., Graminski, G. F. & Armstrong, R. N. *J. biol. Chem.* **266**, 19475-19479 (1991).
- Danielson, U. H. & Mannervik, B. *Biochem. J.* **231**, 263-267 (1985).
- Smith, P. K. et al. *Analyt. Biochem.* **150**, 76-85 (1985).
- Harrington, J. T. & Stastny, P. *J. Immun.* **110**, 752-759 (1973).
- Boyer, T. D. & Kenney, W. C. *D. Zakim Biochem. Pharmacol.* **32**, 297-363 (1983).
- Meyer, J. J. et al. *Biochem. J.* **274**, 409-414 (1991).
- Fjellstedt, T. A., Allen, R. H., Duncan, B. K. & Jakoby, W. B. *J. biol. Chem.* **248**, 3702-3707 (1973).
- Ahmed, A. E. & Anders, M. W. *Biochem. Pharmacol.* **27**, 2021-2025 (1978).
- Kohler-Staub, D. & Leisinger, T. *J. Bact.* **162**, 676-681 (1985).

ACKNOWLEDGEMENTS. We thank T. H. Hostetter for donated rat livers; R. Gunther for surgical assistance; C. Woodward and L. Ellis for technical discussion; and E. Eccleston Jr and the Microchemical Facility, University of Minnesota Medical School for N-terminal sequencing. This work was supported by the NIH.

Exocytotic fusion is activated by Rab3a peptides

Andres F. Oberhauser*, Jonathan R. Monck*, William E. Balch† & Julio M. Fernandez*

* Department of Physiology and Biophysics, Mayo Clinic, Rochester, Minnesota 55905, USA

† Department of Molecular and Cellular Biology, Scripps Research Institute, La Jolla, California 92037, USA

STUDIES of intracellular traffic in yeast and mammalian systems have implicated members of the Rab family of small GTP-binding proteins as regulators of membrane fusion¹⁻¹⁰. We have used the patch clamp technique to measure exocytotic fusion events directly and investigate the role of GTP-binding proteins in regulating exocytosis in mast cells. Intracellular perfusion of mast cells with GTP- γ S is sufficient to trigger complete exocytotic degranulation in the absence of other intracellular messengers¹¹. Here we show that GTP is a potent inhibitor of GTP- γ S-induced degranulation, indicating that sustained activation of a GTP-binding protein is sufficient for membrane fusion. We have found that synthetic oligopeptides, corresponding to part of the effector domain of Rab3a¹², stimulate complete exocytotic degranulation, similar to that induced by GTP- γ S. The response is selective for Rab3a sequence and is strictly dependent on Mg^{2+} and ATP. This suggests that sustained activation of a Rab3 protein causes exocytotic fusion. The peptide response can be accelerated by GDP- β S, suggesting that Rab3a peptides compete with endogenous Rab3 proteins for a binding site on a target effector protein, which causes fusion on activation.

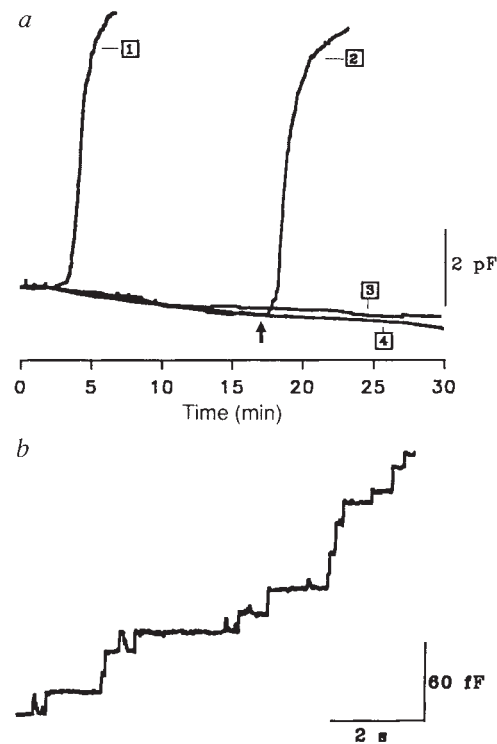
The inclusion of the hydrolysis-resistant GTP analogue, GTP- γ S, in the patch pipette solution causes the complete exocytotic degranulation of mouse mast cells after a delay of about 2 min¹¹

Received 1 July; accepted 1 October 1992.

- Mannervik, B. & Danielson, U. H. *CRC Crit. Rev. Biochem.* **23**, 283-337 (1988).
- Pickett, C. B. & Lu, A. Y. H. *A. Rev. Biochem.* **58**, 743-764 (1989).
- Boyer, T. D. *Hepatology* **9**, 486-489 (1989).

FIG. 1 Exocytotic fusion, measured in patch clamped mast cells requires the sustained activation of a GTP-binding protein. Exocytosis was followed by measuring the cell membrane capacitance using the whole-cell patch clamp technique. The membrane capacitance is proportional to the cell surface area and detects the increases in cell membrane area as secretory granule membrane is incorporated into the cell membrane upon fusion. *a*, GTP- γ S (10 μ M) induces a complete degranulation of a mast cell [1]. In this cell the capacitance increased from an initial value of 3.9 pF to a final value of 13.5 pF as several hundreds of secretory granules (358 for this cell) fuse sequentially with the cell membrane. To demonstrate that the exocytotic response to GTP- γ S does not depend on soluble cytosolic components we delayed the stimulus by using caged-GTP- γ S. [2] Time course of degranulation induced by illuminating DMNPE-caged GTP- γ S with ultraviolet light 17 min (marked by an arrow) after entering the whole-cell configuration. The pipette solution contained 110 μ M of guanosine-5'-(3-thiotriphosphate), 3-S-(1-(4,5-dimethoxy-2-nitrophenyl) ethyl) thioester, trilitium salt (DMNPE-caged GTP- γ S). GDP- β S (40 μ M) was also included to prevent the degranulation induced by contaminant GTP- γ S (estimated to be about 0.1 μ M). The GTP- γ S-induced degranulation is completely inhibited by 0.5 mM GTP [4] ($n=4$), and by 1 mM GDP- β S [3] ($n=5$). Time zero for these recording corresponds to the beginning of the perfusion of the cell with pipette solution (that is on breaking into whole-cell recording mode). In [1] the membrane area starts to expand after a lag period of 2.1 min (mean 1.7 ± 0.1 min, $n=30$) with a rate of 3.3 pF min^{-1} ($2.7 \pm 0.1 \text{ pF min}^{-1}$, $n=30$). The gradual decline in capacitance, evident in all traces before degranulation commences, has been attributed to pinocytosis²⁷. *b*, Step changes in cell membrane capacitance, caused by the exocytotic fusion of individual secretory granules. These steps are part of an expansion of recording [1] taken 4 min after the beginning of cell perfusion with 10 μ M GTP- γ S.

METHODS. Mast cells were prepared from adult C57BL/6J mice of either sex (Jackson Laboratories, Bar Harbor) following a procedure described in detail elsewhere²⁸. The standard extracellular medium used for the experiments was a modified Ringer's solution containing (in mM): 150 NaCl, 10 HEPES, 2.8 KOH, 1.5 NaOH, 1 MgCl₂, 2 CaCl₂ and 25 Glucose (310 mmol kg⁻¹, pH 7.25). The pipette solution contained in mM: 125 K-glutamate, 10 HEPES, 7 MgCl₂, 3 KOH, 0.2 ATP, 9 K₂-EGTA, 1 Ca-EGTA, and various concentrations of guanine nucleotides and synthetic peptides (290 mmol kg⁻¹, pH 7.2). The free Ca²⁺ concentration of these buffers is 30 nM. The Ca²⁺-EGTA buffer was replaced with 50 mM EGTA for some of the experiments with Rab peptides, but had no effect on the results. Fire-polished pipettes were coated



with Sylgard resin and had resistances of 1–3 Mohm when immersed in the external solution. The cell membrane capacitance was measured using the whole-cell mode of the patch clamp technique and a digital phase detector^{29–31}; temperature 22–24 °C. The lag period was defined as the time from the beginning of whole-cell recording mode to the first step increase in capacitance. The rate of degranulation was calculated between 20% and 80% of the final capacitance change. Measurements give mean $\pm$ s.e. (number of cells).

(Fig. 1a). Throughout the degranulation, stepwise increases in the cell surface area can be observed as single secretory granules fuse with the cell membrane (Fig. 1b). In contrast to stimulation by antigens or 48/80 (refs 13, 14), the ability of GTP- γ S to stimulate exocytosis does not wash out, because release of GTP- γ S by ultraviolet photolysis of caged GTP- γ S is still able to induce complete degranulation with similar kinetics after 17 min of intracellular perfusion (Fig. 1a). This time is long enough for more than 99% of freely soluble proteins (<200K) to have diffused out of the cell into the effectively infinite volume of the patch pipette¹⁴. Therefore, the GTP-binding protein activated by GTP- γ S must be close to, or part of, the exocytotic fusion pore machinery and not a soluble component of the cytosol.

The exocytotic response to GTP- γ S is totally abolished by both GDP- β S and GTP, whereas GTP alone has no effect (Fig. 1a). The potent inhibition by GTP suggests that GTP and GTP- γ S compete for binding to the GTP-binding protein and that GTP is hydrolysed before membrane fusion is activated. Thus, sustained activation of a GTP-binding protein is required for membrane fusion.

Rab proteins are members of the Ras superfamily of GTP-binding proteins. The Ras protein has been extensively characterized structurally and biochemically allowing identification of the domains involved in guanine nucleotide binding and hydrolysis, membrane localization and interaction with GTPase-activating protein (GAP) and effector proteins¹⁵. Like Ras, Rab proteins contain structurally conserved regions that are thought to function as the putative GTP-binding and effector domains. Small peptides mimicking specific regions of Ras and Rab proteins have been synthesized and used to probe structure-function relationships^{12,16}. Rab3a peptide is a 16-amino-acid

peptide (sequence Val-Ser-Thr-Val-Gly-Ile-Asp-Phe-Lys-Val-Lys-Thr-Ile-Tyr-Arg-Asn) comprising the native sequence of Rab3a corresponding to amino-acids 33–48 of the Ha-Ras sequence, a sequence implicated in both effector interaction and GAP binding¹⁵. Rab3AL has the same sequence (Val-Ser-Ala-Leu-Gly-Ile-Asp-Phe-Lys-Val-Lys-Thr-Ile-Tyr-Arg-Asn) except for Ala and Leu residues in place of Thr 35 and Val 36 (ref. 12). The analogous substitutions in Ras inhibit the ability of GAP to stimulate the Ras GTPase activity^{16,17}. Recently, Rab3AL and Rab3a peptides have been shown to mimic the effects of GTP- γ S in inhibiting sustained vesicular traffic from the endoplasmic reticulum to the Golgi and between Golgi compartments *in vitro*¹².

Intracellular perfusion of mast cells with either native Rab3a or mutant Rab3AL peptides caused a rapid complete degranulation (Fig. 2a). The peptide was added to the pipette solution in the absence of guanine nucleotides. As with GTP- γ S, individual fusion events can be visualized as step increases in cell membrane area (not shown), which verifies that membrane fusion is occurring. Like GTP- γ S (10 μ M), both Rab3a and Rab3AL peptides ($\sim 100 \mu$ M) fully degranulate mast cells, although the latent period before the degranulation begins is larger with Rab3AL (11.3 ± 0.8 min, $n=21$) or Rab3a (13.2 ± 1.6 min, $n=5$) than with GTP- γ S (1.7 ± 0.1 min, $n=30$). Rab1AL (Ile-Ser-Ala-Leu-Gly-Val-Asp-Phe-Lys-Ile-Arg-Thr-Ile-Glu-Leu-Asp), the analogous 16-mer peptide for the Rab1a protein, only elicits a slow degranulation beginning after about 25 min, whereas another control peptide (Glu-Pro-Thr-Lys-Ala-Asp-Ser-Tyr-Arg-Lys-Lys-Val-Val-Lys-Asp), identical to the effector domain sequence of Ra1 (a non-Rab family small GTP-binding protein), had no effect for more than 40 min (Fig. 2a).

Figure 2b shows that the 5-mer Rab3a (Val-Ser-Thr-Val-Gly), the C-terminal residues of the 16-mer Rab3a and Rab3AL (Val-Ser-Ala-Leu-Gly) also induce complete degranulation of mast cells, although higher concentrations of peptide are required for an equivalent effect. The 5-mer Rab3AL does not include the three basic and single acidic residues in the 16-mer Rab3AL so is clearly unrelated structurally to polycationic mast cell degranulating agents such as mastoparan, substance P, and compound 48/80 (refs 18–20), which are thought to activate heterotrimeric G proteins by direct interaction with the receptor

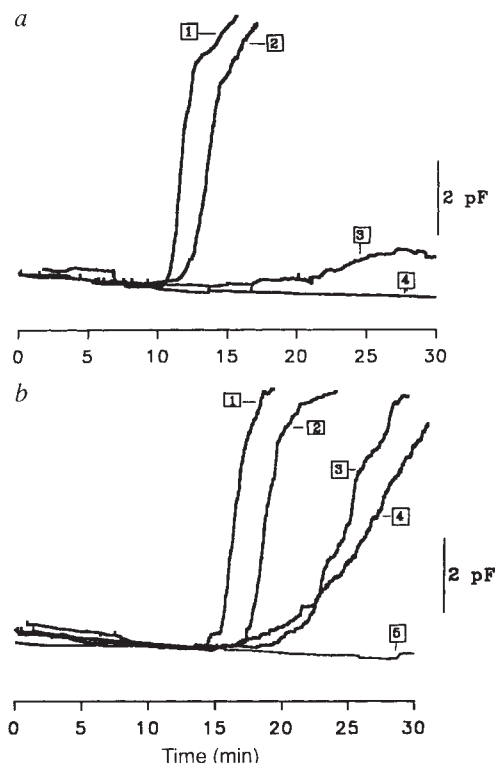


FIG. 2 Selective stimulation of exocytosis by intracellular perfusion with effector domain synthetic peptides. *a*, 16-mer Rab3a (sequence VSTVGIDFKVKTIYRN (single-letter amino-acid code) 125 μ M, [2]) and Rab3AL (sequence VSALGIDFKVKTIYRN 92 μ M, [1]) peptides induce complete exocytotic degranulation. The latencies before onset of the degranulations were 11.3 ± 0.8 min ($n=21$) for Rab3AL peptide and 13.2 ± 1.6 min ($n=5$) for Rab3a peptide. The mean rates of degranulation were 1.3 ± 0.2 pF min $^{-1}$ ($n=14$) and 1.2 ± 0.4 pF min $^{-1}$ ($n=5$), respectively. Rab1AL (ISALGVDFKIR-TIELD, the analogous mutant sequence for Rab1a, 100 μ M) only induced a slow exocytotic response after a mean delay of 22.1 ± 3.0 min [4] (three cells). Ral peptide (sequence EPTKADSYRKKVVD, the align sequence for Ral, a non-Rab family small G protein) failed to activate exocytosis [5] (140 μ M; two cells). In contrast, the 16-mer Rab3a and Rab3AL peptides were effective in activating degranulation in all cells tried (42 cells). Non-specific effects caused by Ca^{2+} leak through the gigaseal or cell membrane, or by artifactual ion channel activity due to the peptide, are extremely unlikely to explain the stimulation of exocytosis because the experiments with peptides used 10 or 50 mM EGTA in the patch pipette solution. Moreover, the membrane conductance was continuously monitored during the degranulation and no channel-like activities were observed. *b*, The 5-mer native Rab3a (VSTVG 390 μ M, [1]) and mutant Rab3AL (VSALG 400 μ M, [2]) peptides also induce a complete exocytotic degranulation, like the 16-mer but higher concentrations of peptide are required. In contrast, the native 5-mer Rab1a (ISTIG 400 μ M, [3]) and mutant Rab1AL (ISALG 400 μ M, [4]) peptides were less effective. The latencies for the Rab3a and Rab1a native sequences were 11.9 ± 1.3 min ($n=8$) and 16.2 ± 2.1 min ($n=8$), and the mean rates were 1.4 ± 0.2 pF min $^{-1}$ ($n=4$) and 0.92 ± 0.15 pF min $^{-1}$ ($n=5$), respectively. For the Rab3AL and Rab1AL mutant sequences the latencies were 15.4 ± 0.7 min ($n=8$) and 19.1 ± 3.6 min ($n=8$), and the mean rates were 1.8 ± 0.4 pF min $^{-1}$ ($n=8$) and 0.96 ± 0.21 pF min $^{-1}$ ($n=5$), respectively. It is clear that the delays are much longer and the rates slower for the Rab1a peptides than with the corresponding Rab3a peptides. The 5-mer Ras peptide (DPTIE 375 μ M, [5]) was unable to stimulate exocytosis. The peptides used in these studies had an open N-terminus and a closed amidated C terminus.

binding domain¹⁹. As with the 16-mer peptides, 5-mer Rab1AL (Ile-Ser-Ala-Leu-Gly) and native Rab1a (Ile-Ser-Thr-Ile-Gly) peptides were less effective stimulators of exocytosis than the Rab3a peptides, whereas the 5-mer (Asp-Pro-Thr-Ile-Glu) corresponding to the Ras effector domain was unable to stimulate exocytosis (Fig. 2b). It is surprising that the 5-mer peptides, that have much less in common with the effector loop, are still effective. Although we cannot rule out that their mode of action is different from that of the 16-mers, the similar selectivities suggests the same mechanism.

At first sight, the ability of GTP- γ S and Rab3a peptides to stimulate exocytotic fusion in the mast cell contrasts with their inhibitory action on vesicular transport^{4–9,12}. But a sustained interaction of GTP- γ S-Rab or Rab3a peptides with their effector proteins may cause fusion initially, but then prevent the recycling of components required for subsequent rounds of vesicle fusion, thus inhibiting sustained vesicular traffic^{3,6}.

One possible explanation for the action of the Rab3a peptides is that they bind to a Rab GAP protein and prevent GTP hydrolysis by endogenous Rab proteins, which will then remain constitutively active and trigger exocytosis. This interpretation predicts that GTP is necessary for the response to Rab3a peptides. Thus, exogenously added GTP should potentiate the

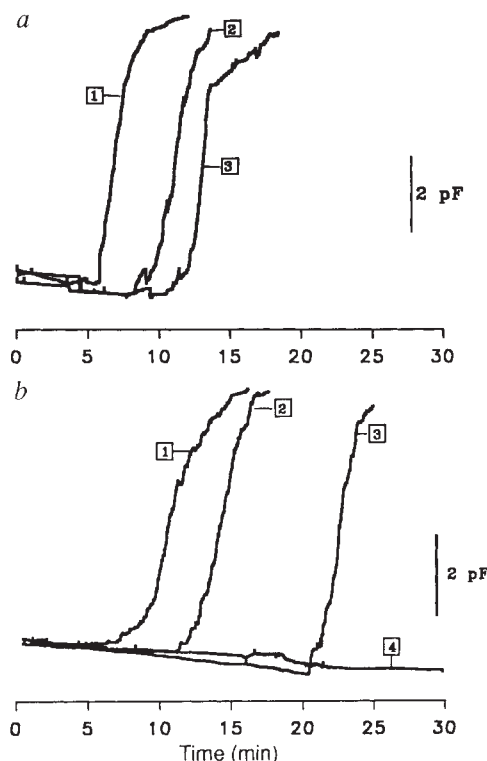


FIG. 3 Modulation of Rab3a peptide responses by guanine and adenine nucleotides. *a*, The 16-mer Rab3AL peptide (92 μ M) induces a complete exocytotic degranulation [2]. The exocytotic response is accelerated in the presence of 2 mM GDP- β S [1] but not with 2 mM GTP [3]. The latencies before onset of the degranulation were 11.3 ± 0.8 min ($n=21$) for Rab3AL peptide alone, 10.8 ± 1.6 min ($n=5$) for peptide with GDP- β S, and 7.5 ± 1.2 min ($n=11$) for peptide with GDP- β S. The mean rates of degranulation were 1.3 ± 0.2 pF min $^{-1}$ ($n=14$), 1.0 ± 0.4 pF min $^{-1}$ ($n=3$) and 1.9 ± 0.2 pF min $^{-1}$ ($n=8$), respectively. *b*, The 5-mer Rab3AL peptide (sequence VSALG) induces a complete exocytotic degranulation at 400 μ M [2] and 200 μ M [3]. The response to 400 μ M of the 5-mer peptide is accelerated by 1 mM GDP- β S [1]. The peptide induced degranulation is completely inhibited by removing Mg^{2+} and ATP from the internal solution [4] ($n=3$). The latencies before onset of the degranulation were 26.1 ± 4.0 min ($n=3$) for the peptide at 200 μ M, 15.4 ± 0.7 min ($n=8$) for the peptide at 400 μ M and 10.0 ± 2.5 min ($n=3$) for 400 μ M peptide with GDP- β S. The mean rates of degranulation were 1.0 pF min $^{-1}$ ($n=1$), 1.8 ± 0.4 pF min $^{-1}$ ($n=8$) and 1.1 ± 0.1 pF min $^{-1}$ ($n=3$), respectively. In [4] the pipette solution contained 14 mM KCl instead of 7 mM MgCl_2 to maintain the chloride concentration.

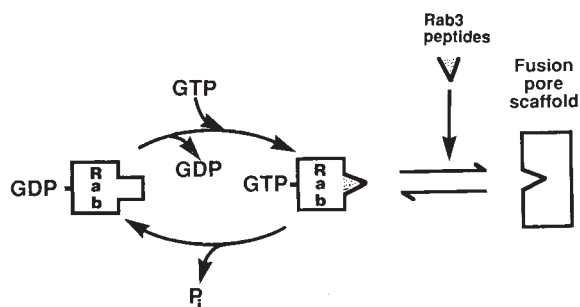


FIG. 4 A model to describe the regulation of exocytosis in mast cells by Rab proteins. Rab is activated by binding of GTP and the active GTP-Rab complex binds to an unknown effector protein, which is part of the exocytotic fusion pore scaffold. GTP is rapidly hydrolysed and is unable to generate the sustained stimulus necessary for exocytosis. GTP- γ S, which forms a stable GTP- γ S-Rab complex, causes a sustained Rab activation. Rab3 peptides directly bind the effector protein and are able to provide a sustained effector activation. The requirement for a sustained stimulus can be explained if there is a second condition necessary for exocytosis (not shown) and this condition is met only infrequently in patch-clamped cells (that is without a physiological stimulus). This extra condition is necessary to explain why GTP-Rab does not activate. Candidates for this condition include Ca^{2+} -binding proteins, heterotrimeric G-proteins, protein kinase C and phosphoprotein phosphatases. Physiological stimuli would increase the occurrence of the second condition and/or increase the lifetime of the GTP-Rab complex.

response and GDP- β S should inhibit by locking endogenous Rab protein in the inactive state. In contrast to these expectations the Rab3AL response is not potentiated by the presence of GTP in the patch pipette, and GDP- β S does not inhibit at concentrations as high as 4 mM (Fig. 3a), in striking contrast to the strong inhibition of the GTP- γ S-induced exocytosis (Fig. 1a). Surprisingly, GDP- β S shortened the delay before exocytosis began, the lag being 11.3 ± 0.8 ($n = 21$) min without and 7.5 ± 1.2 ($n = 11$) min with GDP- β S. Figure 3b shows that the response to 5-mer Rab3AL is also accelerated by GDP- β S. These results provide strong evidence that the exocytotic response to Rab3a peptides does not result from changes in the GTPase activity of the putative GTP-binding protein and, instead, indicates that the peptide competes with an endogenous GTP-binding protein for binding to a target effector protein, which is able to activate exocytosis.

Figure 3b also shows that higher concentrations of 5-mer Rab3AL peptide, as with higher GTP- γ S concentrations, reduce the lag before onset of exocytosis and that the peptide response strictly depends on the presence of Mg^{2+} and ATP (Fig. 3). That the peptide response can be modulated both by adenine and by guanine nucleotide eliminates the possibility that nonspecific membrane perturbing effects explain the action of the peptides.

The ability of Rab3a peptides to stimulate mast cell degranulation suggests that a small GTP-binding protein of the Rab3 subfamily might be G_e , the putative GTP-binding protein thought to regulate exocytosis²¹. Rab3a, one member of this subfamily has been implicated in neurotransmitter secretion because it has been found only in secretory cells²², is localized to synaptic vesicles^{23–25}, and undergoes cycles of dissociation and reassociation with synaptic vesicles during exocytosis¹⁰. It is possible that different members of the Rab3 subfamily (Rab3a–d) regulate exocytosis in different cells.

A simple model to explain the role of a Rab3 protein in the regulation of exocytosis is shown in Fig. 4. In the presence of GTP, the Rab protein is only activated transiently because of a high GTPase activity, presumably due to interaction with a GAP protein. Because GTP alone does not activate, the lifetime of the GTP-Rab3 complex must be too short for activation of other molecular events that are also required for exocytosis. GTP- γ S is hydrolysis resistant, leading to sustained activation of the Rab3 protein and subsequent activation of the effector pro-

tein(s). This effector protein could be part of the scaffold of proteins that has recently been proposed to direct the formation of the exocytotic fusion pore²⁶. We do not know the molecular mechanism by which Rab3a peptides act, but a plausible hypothesis is that the peptides directly activate the effector proteins. The ability of GDP- β S to shorten the lag before onset of exocytosis supports the interpretation that the Rab3 peptides compete with endogenous GTP-bound Rab3 proteins for binding to the effector. In the presence of GDP- β S, the endogenous Rab3 will be inactivated and the competition for effector binding will be relieved, allowing exocytosis to proceed earlier.

By directly measuring exocytotic fusion in patch clamped mast cells we have provided evidence that a sustained activation of a Rab3 protein is sufficient to cause exocytosis. Because the putative Rab3 stimulus is effective after prolonged washout of the cytosol, the effector must be close to, or part of, the scaffold of proteins that causes exocytotic fusion²⁶.

Note added in proof: Since the submission of this paper, the ability of Rab3a peptides to stimulate exocytosis was reported for permeabilized pancreatic acinar cells³² and adrenal chromaffin cells³³.

Received 3 August; accepted 6 October 1992.

1. Bourne, H. R. *Cell* **53**, 669–671 (1988).
2. Balch, W. E. *J. Biol. Chem.* **264**, 16965–16968 (1989).
3. Walworth, N. C., Goud, B., Kabcenell, A. K. & Novick, P. J. *EMBO J.* **8**, 1685–1693 (1989).
4. Baker, D., Hicke, L., Rexach, M., Schleyer, M. & Schekman, R. *Cell* **54**, 335–344 (1988).
5. Beckers, C. J. M. & Balch, W. E. *J. Cell Biol.* **108**, 1245–1256 (1989).
6. Rexach, M. F. & Schekman, R. W. *J. Cell Biol.* **114**, 219–229 (1991).
7. Melancon, P. et al. *Cell* **51**, 1053–1062 (1987).
8. Mayorga, L. S., Diaz, R. & Stahl, P. D. *Science* **244**, 1475–1477 (1989).
9. Mayorga, L. S., Diaz, R., Colombo, M. I. & Stahl, P. D. *Cell Reg.* **1**, 113–124 (1989).
10. Fischer von Mollard, G., Südhof, T. C. & Jahn, R. *Nature* **349**, 79–81 (1991).
11. Fernandez, J. M., Neher, E. & Comperts, B. D. *Nature* **312**, 453–455 (1984).
12. Plutner, H., Schwaninger, R., Pind, S. & Balch, W. E. *EMBO J.* **9**, 2375–2383 (1990).
13. Penner, R., Pusch, M. & Neher, E. *Biosci. Rep.* **7**, 313–321 (1987).
14. Pusch, M. & Neher, E. *Pflügers Arch.* **411**, 204–211 (1988).
15. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **349**, 117–127 (1991).
16. Schaber, M. D. et al. *Protein Struct. Funct. Genet.* **6**, 306–315 (1989).
17. Vogel, U. S. et al. *Nature* **335**, 90–93 (1988).
18. Lindau, M. & Nüsse, O. *FEBS Lett.* **222**, 317–321 (1987).
19. Mousli, M. et al. *J. Pharmac. exp. Ther.* **250**, 329–335 (1989).
20. Mousli, M., Bueb, J.-L., Bronner, C., Rouot, B. & Landry, Y. *Trends pharmac. Sci.* **11**, 358–362 (1990).
21. Gomperts, B. D. A. *Rev. Physiol.* **52**, 591–606 (1990).
22. Mizoguchi, A., Kim, S., Ueda, T. & Takai, Y. *Biochem. biophys. Res. Commun.* **162**, 1438–1445 (1989).
23. Mizoguchi, A. et al. *J. Biol. Chem.* **265**, 11872–11879 (1990).
24. Fischer v. Mollard, G. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1988–1992 (1990).
25. Darchen, F. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5692–5696 (1990).
26. Monck, J. R. & Fernandez, J. M. *J. Cell Biol.* (in the press).
27. Almers, W. & Neher, E. *J. Physiol.* **386**, 205–217 (1987).
28. Alvarez, de Toledo, G. & Fernandez, J. M. *J. Cell Biol.* **110**, 1033–1039 (1990).
29. Joshi, C. & Fernandez, J. M. *Biophys. J.* **53**, 885–892 (1988).
30. Fidler, N. & Fernandez, J. M. *Biophys. J.* **56**, 1153–1162 (1989).
31. Monck, J. R., Alvarez de Toledo, G. & Fernandez, J. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7804–7808 (1990).
32. Padfield, P. J., Balch, W. E. & Jamieson, J. D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1656–1660 (1992).
33. Senyshyn, J., Balch, W. E. & Holz, R. W. *FEBS Lett.* **309**, 41–46 (1992).

ACKNOWLEDGEMENTS. We thank T. Südhof and P. Novick for reading of the manuscript and C. Camrud for secretarial assistance.

Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*

M. Alejandra Mandel, Cindy Gustafson-Brown, Beth Savidge & Martin F. Yanofsky*

Department of Biology and Center for Molecular Genetics, University of California at San Diego, La Jolla, California 92093-0116, USA

* To whom correspondence should be addressed

THE first step in flower development is the transition of an inflorescence meristem into a floral meristem. Each floral meristem differentiates into a flower consisting of four organ types that occupy precisely defined positions within four concentric whorls.

Genetic studies in *Arabidopsis thaliana* and *Antirrhinum majus* have identified early-acting genes that determine the identity of the floral meristem, and late-acting genes that determine floral organ identity¹⁻⁵. In *Arabidopsis*, at least two genes, *APETALA1* and *LEAFY*, are required for the transition of an inflorescence meristem into a floral meristem¹. We have cloned the *APETALA1* gene and here we show that it encodes a putative transcription factor that contains a MADS-domain². *APETALA1* RNA is uniformly expressed in young flower primordia, and later becomes localized to sepals and petals. Our results suggest that *APETALA1* acts locally to specify the identity of the floral meristem, and to determine sepal and petal development.

Mutations in the *APETALA1* (*API*) gene⁶ result in a conversion of sepals in the first whorl into leaf-like organs, which often

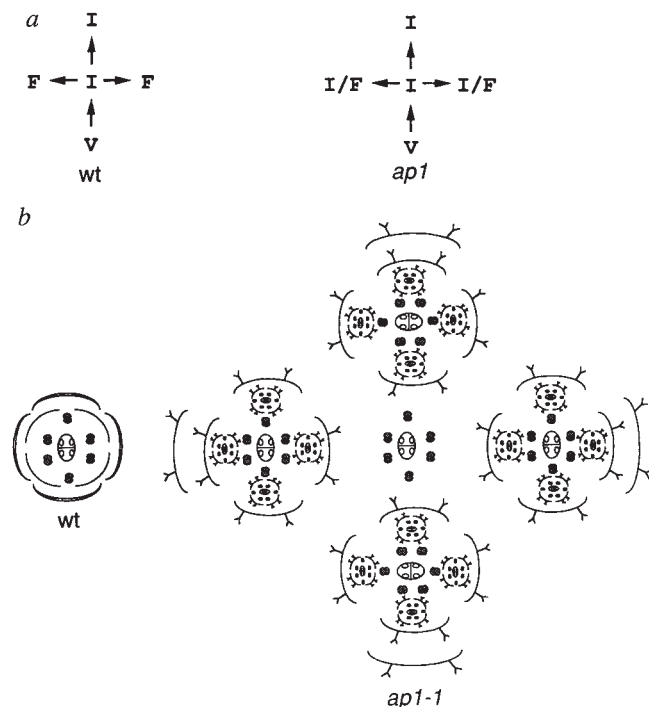
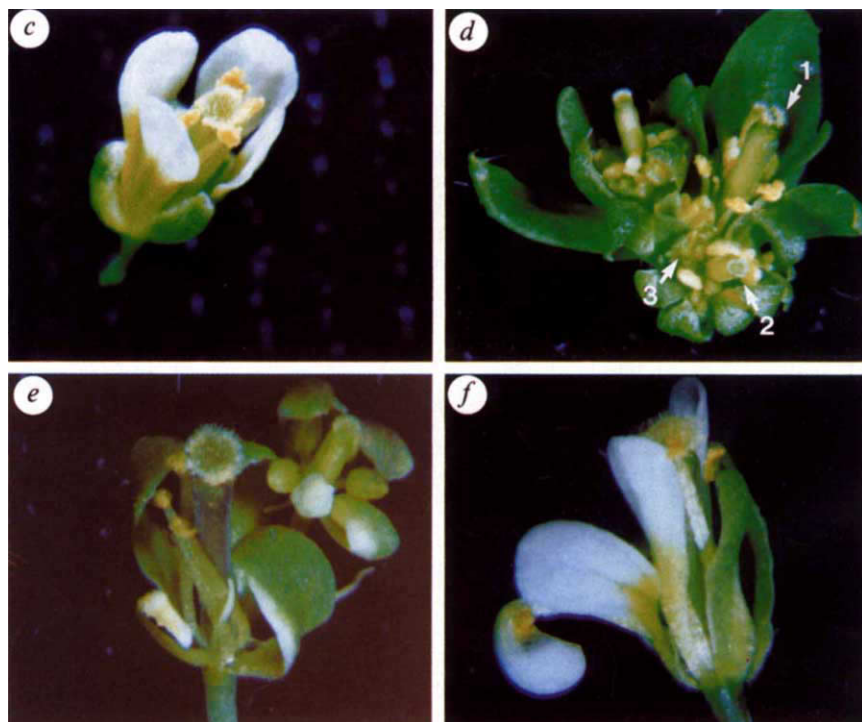


FIG. 1 Comparison of wild-type (wt) and *ap1* mutant plants. *a*, Meristems produced by wild-type and *ap1* plants. V, vegetative meristem gives rise to roots and leaves; I, inflorescence meristem gives rise to floral meristems; F, floral meristem gives rise to floral organs. A partial conversion of inflorescence to floral meristems occurs in *ap1* mutants (I/F). *b*, Diagram of wild-type and *ap1* flowers. Wild-type *Arabidopsis* flowers consist of four sepals in the first whorl (outermost), four petals in the second, six stamens in the third, and two fused carpels in the fourth (innermost). Stellate trichomes (Y) develop on the first whorl leaf-like organs of *ap1-1* flowers. *c-f*, Photographs of *Arabidopsis* flowers. *c*, Wild-type; *d*, *ap1-1*. The numbered arrows indicate the positions of the primary (1), secondary (2) and tertiary (3) flowers. *e*, *ap1-2* flower where mosaic organs, and a single axillary flower are evident. *f*, *ap1-3* flower where the second whorl is often occupied by petal-stamen mosaics of filamentous structures.

subtend secondary flowers in their axils⁷. This pattern may be repeated, such that the axillary secondary flowers produce tertiary flowers in the axils of their first whorl organs. The development of secondary flowers can be interpreted as a partial conversion of a floral meristem into an inflorescence meristem (Fig. 1*a*). Second whorl organs fail to develop in *ap1-1* mutant flowers⁷, although petaloid organs occasionally develop in higher order *ap1-1* flowers. The stamens and carpels that occupy the third and fourth whorls, respectively, appear to be normal (Fig. 1*b*). Therefore, genetic studies indicate that *API* is required for determining the identity of the floral meristem, and for proper development of first and second whorl organs.

Several floral homeotic genes have been isolated, including *AGAMOUS* (*AG*) and *APETALA3* (*AP3*) in *Arabidopsis*, and *DEFICIENS A* (*DEF A*) and *SQUAMOSA* (*SQUA*) in *Antirrhinum*^{2,8-11}. Their gene products contain a DNA-binding domain, called the MADS-domain², which they share with the transcription factors MCM1 and SRF^{2,12-16}. Six additional MADS-box-containing genes, designated *AGL1-6* (*AGAMOUS* like) have also been isolated from *Arabidopsis*¹⁷.

We have isolated a new MADS-box-containing gene, *AGL7* (Fig. 2). *AGL7* maps at or near the *ap1-1* mutation on chromosome 1 on a restriction fragment length polymorphism (RFLP) map (ref. 18; Fig. 2). To determine whether *AGL7* is the *API* gene, a 9.5 kilobase pair (kbp) genomic restriction DNA fragment containing the wild-type *AGL7* gene was introduced into *ap1-1* homozygous mutant plants. Several independently regenerated plants had flowers with a wild-type phenotype, demonstrating that this DNA fragment is sufficient to rescue the mutated gene (data not shown). Nucleotide sequence analysis of strong (*ap1-1*; Fig. 1*d*), intermediate (*ap1-2*; Fig. 1*e*) and weak (*ap1-3*; Fig. 1*f*) *ap1* alleles correlates the type of mutation with the corresponding severity of mutant phenotype (Fig. 2*a*). On the basis of the RFLP analysis, rescue of the mutant phenotype, and sequence analysis of three independently isolated mutant alleles, we conclude that *AGL7* is the *API* gene. The *API* complementary DNA encodes a putative protein of *M_r* 30,000 (30K) (Fig. 2*a*) that contains the highly conserved MADS-domain (Fig. 2*b*). A second domain called the K-domain¹⁷, a region with structural similarity to the coiled-coil domain of keratins¹⁹, is also present.



CTT
TCCAATTGGTTCATACCAAGTCTGAGCTCTTCTTTATATCTCTTGTAGTTTCTTATT
GGGGTCTTTGTTTGGTTCTTTAGAGTAAGAAGTTTCTTAAAAAGGATCAAAA
ATGGGAAGGGTAGGTTCAATTGAAGAGGATAGAGAACAAGATCAATAGACAAGTGACA 60
M G R G R V Q L K R I E N K I N R Q V T 20
▲ (ap1-2)
TTCTCGAAAAGAAGAGCTGGTCTTTTGAAGAAAGCTCATGAGATCTCTGTTCTCTGTGAT 120
F S K R R A G L L K K A H E I S V L C D 40
D
▲ (Col)
GCTGAAGTGTCTTGTGTCTTCTCCATAAGGGGAACTCTTCAATCTCCACTGAT 180
A E V A L V V F S H K G K L F E Y S T D 60
▼
TCTGTATGGAGAAGATACTTGAACGCTATGAGAGTACTCTTACGCCGAAAGACAGCTT 240
S C M E K I L E R Y E R Y S Y A E R Q L 80
▲
ATTGCACCTGAGTCCGACGCTCAATAACAAGTGGTCGATGGAGTATAACAGGCTTAAGGCT 300
I A P E S D V N T N W S M E Y N R L K A 100
▼ (ap1-1)
AAGATTGAGCTTTTGGAGAGAAACAGGAGGATTATCTTGGGGAAGACTTGAAGCAATG 360
K I E L L E R N Q R H Y L G E D L Q A M 120
AGCCCTAAGAGCTTCAGAATCTGGAGCAGCAGCTTGACACTGCTCTTAAGCACATCCGC 420
S P K E L Q N L E Q Q L D T A L K H I R 140
▼
ACTAGAAAAACCACTTATGTACGAGTCCATCAATGAGCTCCAAAAAAGGAGAGGCC 480
T R K N Q L M Y E S I N E L Q K K E K A 160
▼
ATACAGGAGAAAACAGCATGCTTTCTAAACAGATCAAGGAGAGGAAAAATCTTAGG 540
I C Q E Q N S M L G Q I K E R E K I L R 180
▼
GCTCAACAGGAGCAGTGGGATCAGCAGAACCAAGGCCACAATATGCTCTCCCCCTGCGCA 600
A Q Q E Q W D Q Q N Q G H N M P P P L P 200
CGCAGCAGCACCAATCCAGCATCTTACATGCTCTCTCATCAGCATCTCTTTTCTC 660
P Q Q H Q I Q H P Y M L S H Q P S P F L 220
▼
AACATGGGTGGTCTGTATCAAGAAGATGATCCAATGGCAATGAGGAGGAATGATCTCGAA 720
N M G G L Y Q E D D P M A M R R N D L E 240
CTGACTCTTGAACCGTTTACAACCTGCAACCTTGGCTGCTTCGCCGATGAAGCATTTCC 780
L T L E P V Y N C N L G C F A A * 256
ATATATATATTGTAATCGTCAACAATAAAACAGTTTGGCCACATACATATAAATAGTGG 840
CTAGGCTCTTTTCAATCAATTAATATATTCTTGGCAATGTTTCGATGTTCTTATATCATCA 900
C (Col)
TATATAAATAGCAGGCTCTCTTCTTTTGT (A) n 933

AP1 2 GGRVQLKRIENKINRQVTFSKRRAGLLKKAHEISVLCDAEVALVVFSHKGLFEY
AGL2 2 *****E*****A*****N*****Y*L*****II*NR***Y*F
AGL4 2 *****E*****A*****N*****Y*L*****S*I*NR***Y*F
AGL6 2 *****EM*****N*****Y*L*****II*TR***Y*F
AGL3 2 ***K*E*****A*****N*****Y*L*****I*LI*NR***Y*F
AG 51 ***KIEI*****TT*****C*****N*****Y*L*****I*SR*Y**
AGL1 17 ***KIEI*****TT*****C*****N*****Y*L*****I*TR*Y**
AGL5 17 ***KIEI*****TT*****C*****N*****Y*L*****I*TR*Y**
AP3 2 A**KI*****QT*****N*****LT*****R*SIIM*SSN**H**
SQUA 2 ***K*****G*****L*****I*****N*****
DEF A 2 A**KI*****QT*****Y*****N*****F*****L*****K*SIIMI*STQ**H**
SRF 142 **VKIKMEF*D*LR*YT*****KT*IM***Y*L*T*GTQ*L*L*A*ET*HVYTF
MCM1 17 E*RKIEI*F*****TR*H*****F**KH*IM***Y*L*TT*GTQ*L*L*V*ET*LVYTF
ARG80 79 T*RKQPIRY*****TR*H*****H*IM***Y*L*TT*GTQ*L*L*V*ET*LVYTF

AP1 RNA is preferentially expressed in flowers, as no hybridizing RNA was detected in roots, stems or leaves (Fig. 3b). To determine the temporal and spatial expression pattern of *AP1* RNA in developing flowers, *in situ* hybridizations were done on tissue sections of wild-type inflorescences. The earliest expression of *AP1* is in young flower primordia when they initially arise on the flanks of the inflorescence meristem, and expression increases as the flower primordia increase in size (Fig. 4a and b). During this period, *AP1* RNA is uniformly expressed throughout the floral meristem, but not in the inflorescence meristem (Fig. 4b). At later stages of development, expression disappears in the cells that will give rise to stamens and carpels, and *AP1* RNA is confined to sepals and petals (Fig. 4a, c, and d). The loss of *AP1* expression in the cells that will give rise to stamens and carpels coincides with the onset of *AG* expression in this same region (compare Fig 4c and e). In addition, *AP1* RNA is expressed in the pedicel at all stages of flower development.

To determine possible interactions between *AP1* and other floral homeotic genes, we analysed *AP1* expression in flowers

FIG. 2 a, Nucleotide and deduced amino-acid sequences (single-letter code) of *AP1* cDNA. ▼, position of introns as deduced from a comparison between genomic and cDNA sequences from Landsberg *erecta*. Position of mutations in *ap1* alleles are shown above the nucleotide sequence and the amino-acid change generated by the *ap1-2* mutation is shown below the amino-acid sequence. Mutations in *ap1-1* and *ap1-3* alleles occur at splice acceptor sites of the third and fifth introns, respectively. The *ap1-1* mutation is equivalent to the *ag-1* mutation⁹, and the *ap1-2* mutation is equivalent to the *nicotianoides* allele of *DEF A*¹⁶. Alterations in the Columbia ecotype are shown above the nucleotide sequence. b, Comparison of MADS-domain regions. Alignment of AP1, AGL1-6, AG, SQUA, DEF A, AP3, SRF, MCM1, and ARG80. The asterisks indicate identity to AP1. A putative phosphorylation site is indicated by a thin line above the AP1 sequence. Numbers on the left indicate the position of the first amino acid shown.

METHODS. A genomic cosmid library of Columbia DNA was screened at reduced stringency with a 550 bp *EcoRI* fragment of the *AGL3* clone¹⁸. This resulted in the isolation of a new gene, *AGL8* (M.A.M. and M.F.Y., unpublished results), which was used to screen a library of Columbia cDNA in λ -YES²⁰. As a result of this screening, a cDNA for a new *AGL* gene, *AGL7*, was isolated using the same hybridization conditions¹⁷. This new clone was placed on the RFLP map of ref. 18 using an *XbaI* RFLP between Columbia and Landsberg ecotypes, and found to cosegregate with *ap1-1* in 204 meiotic products scored. Additional *AGL7*-cDNA clones of Landsberg were isolated from a cDNA library in λ -ZAP. Genomic libraries of wild-type and mutants, which were all from the Landsberg ecotype, were prepared from DNA extracted from mature plants. DNA was completely digested with *EcoRI*, ligated to EMBL3/*EcoRI* arms and packaged *in vitro* using Gigapack Gold extracts (Stratagene). The libraries were screened with the entire *AGL7* clone using high-stringency conditions. All positive clones were subcloned into pGEM7Zf(+) and double-strand sequenced using the USB Sequenase Kit. Sequence analyses were done using the MacVector program from IBI.

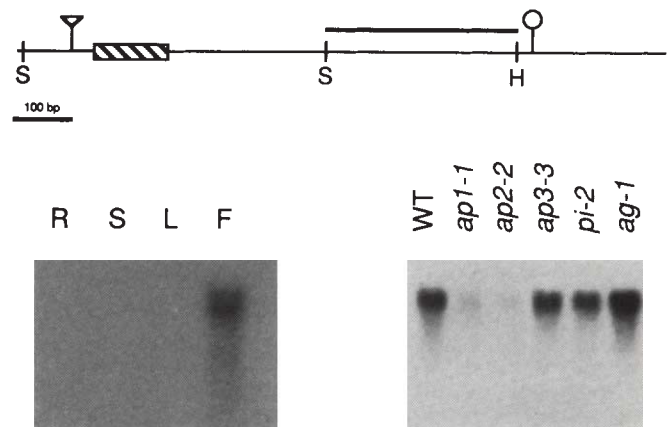


FIG. 3 RNA blots. a, Map of the longest *AP1* cDNA. The symbols ▼ and ○ indicate the positions of translation initiation and termination codons, respectively. The hatched box indicates the position of the MADS-box. The bar above represents the *SacI*-*HincII* region used as a probe for all the expression analyses. A cDNA fragment containing only a portion of the 3' end of the gene was used as a probe to avoid cross-hybridization due to the MADS-box. b, RNA blot of wild-type tissues. RNAs from roots (R), stems (S), leaves (L) and flowers up to stage 12 (F) from *Arabidopsis* probed with the *AP1*-specific probe. c, RNA blot of wild-type and mutant *Arabidopsis* flowers.

METHODS. Total RNAs were isolated as described²², size fractionated by gel electrophoresis on formaldehyde agarose gels, transferred to Hybond N nylon membranes (Amersham) and hybridized according to standard procedures with the ³²P-radiolabelled *AP1*-specific probe described above. Control probes demonstrated that roughly equal amounts of RNA were present in each lane.

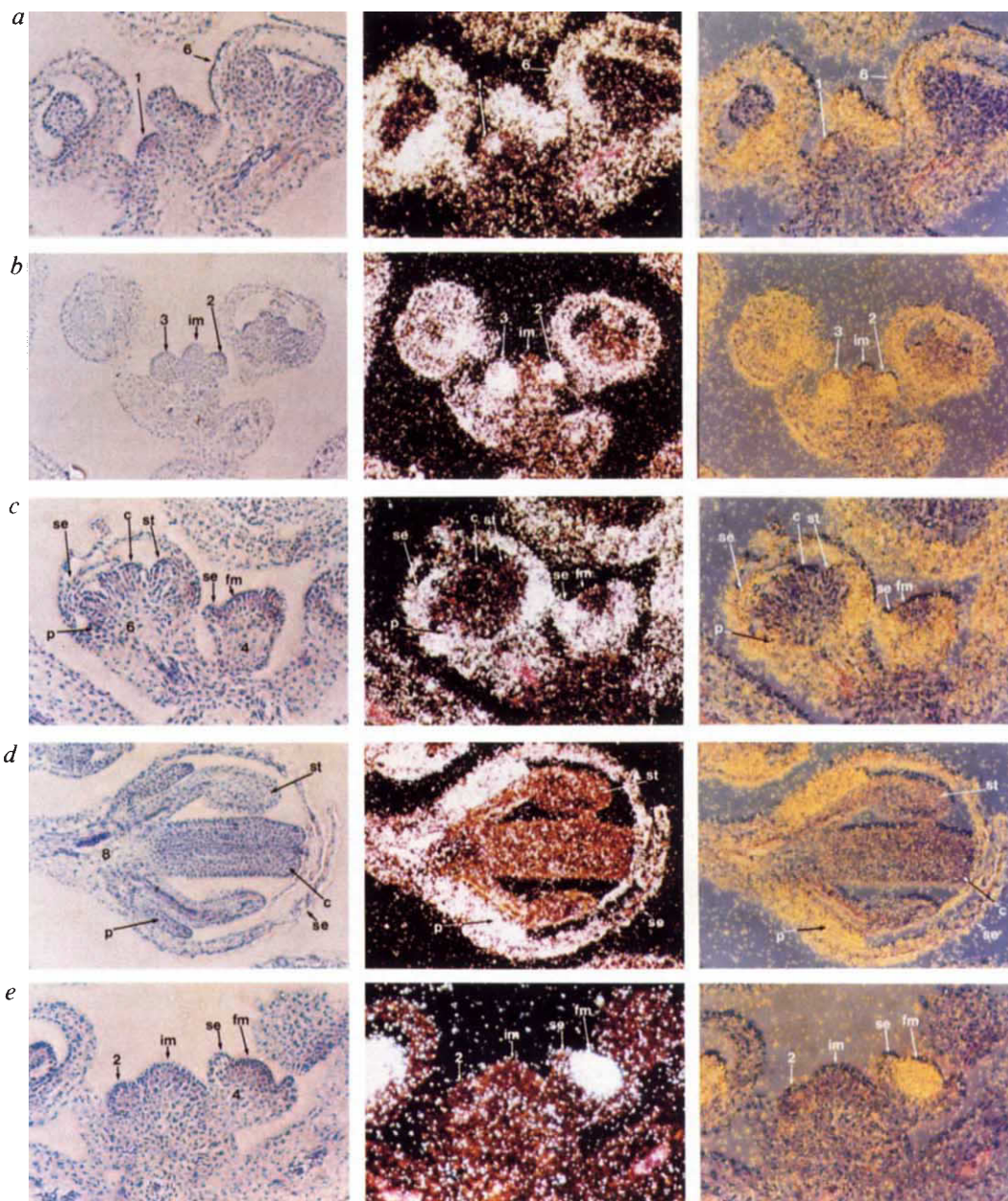
of several homeotic mutants. *AP1* is expressed at similar levels in wild-type, *ap3-3*, and *pistillata-2* (*pi-2*) flowers. In contrast, only low levels of *AP1* RNAs were detected in *ap1-1* or *apetala2-2* (*ap2-2*) flowers, and *AP1* RNA levels are higher in *ag-1* flowers (Fig. 3c). These observations suggest that *AG* is a negative regulator and/or *AP2* is a positive regulator of *AP1* expression. The opposing effects of *AG* and *AP2* are explained in part by the observation that these two genes are mutually antagonistic⁵. We have recently shown that *AP1* is expressed in *ap2* mutant flower primordia, and that *AP1* is expressed in all whorls throughout flower development in *ag* mutants (C.G.-B.

and M.F.Y., manuscript in preparation). Taken together, these data indicate that *AP2* is not required for *AP1* expression and that *AG* negatively regulates *AP1* RNA accumulation.

Genetic studies suggest that *AP1* acts together with *LEAFY* (*LFY*) to determine the identity of the floral meristem¹. The *lfy* mutant displays a complete transformation of early-arising flowers and a partial transformation of late-arising flowers into inflorescence shoots¹. The partial transformation of later arising flowers is strongly enhanced in *ap1 lfy* double mutants, with most of the organs of the double mutant flowers having leaf-like characteristics. This has led to the suggestion that *AP1* and *LFY*

FIG. 4 Expression of *AP1*

RNA in wild-type flowers. *a-d*, Series of photographs of longitudinal sections of wild-type flowers hybridized with an *AP1* antisense mRNA probe. *e*, Longitudinal section of a floral meristem hybridized with an *AG* antisense mRNA probe. The numbered arrows indicate the stage of flower primordia²²; im, inflorescence meristem; fm, floral meristem; se, sepal; p, petal; st, stamen; c, carpel. From left to right, photographs of bright field, dark field, and bright/dark field double exposures respectively; a yellow filter was used for the dark field in the double exposures, causing the signal to appear yellow. The pattern of *AP1* expression was determined by *in situ* hybridization of floral tissues at different stages of development. The earliest expression of *AP1* is observed in young flower primordia (*a*) when they initially arise on the flanks of the inflorescence meristem (stage 1; as defined in ref. 22), and expression increases as the flower primordium increases in size during stage 2 (*b*). During stages 1 and 2, *AP1* RNA is uniformly expressed throughout the flower primordium. In contrast, no *AP1* RNA was detected in the inflorescence meristem (*b*). By the time the sepal primordia arise (stage 3), *AP1* expression has abated in the central dome (the cells that will give rise to the inner whorl organs; *b* and *c*). No *AP1* RNA is detected in the stamen or carpel primordia, or in the cells that will give rise to these organ primordia from stage 3 onwards. During stage 4, when the sepal primordia elongate, *AP1* expression is detected in sepal primordia but is absent in the centre of the developing flower (*c*). By stage 6, when the petal primordia are clearly evident, *AP1* RNA is observed both in sepal and petal primordia (*a* and *c*), and continues to be localized to the sepal and petal primordia during later stages of development (*d*). In contrast to *AP1*, which is expressed in young flower primordia, expression of the organ identity genes *AG* and *AP3* is first observed in stage 3 flowers^{11,23}. The onset of *AG* expression in cells that will give rise to the inner two whorls coincides with the loss of *AP1*



expression in this same region (compare *c* and *e*).

METHODS. ³⁵S-labelled antisense mRNA was synthesized from the *AP1* template shown in Fig. 3a, and hybridized to 8 µm sections of flowers of different stages. Hybridization conditions were the same as previously described²³ with minor modifications. Probes were used at a final concentration of 5 × 10⁷ d.p.m. ml⁻¹. Slides were exposed for two weeks. A control (sense) probe was transcribed in the reverse orientation on the same template. The *AG* antisense probe was synthesized from a *Hind*III-digested pCIT565 template as previously described²³.

act synergistically, and that they may have overlapping functions in determining the identity of the floral meristem¹. Like *API*, *LFY* is expressed in young flower primordia¹. The overlapping expression patterns of *API* and *LFY* support the notion that the two gene products interact early in flower development to determine the identity of the floral meristem.

The deduced amino-acid sequence of *API* shares 68% identical amino-acid residues with the deduced amino-acid sequence of *SQUA* in *Antirrhinum*¹⁰, suggesting that *SQUA* is the

homologue of *API*. A similar level of sequence conservation was observed between *LFY* and its *Antirrhinum* counterpart *FLORICAULA* (*FLO*)^{1,4}. But mutations in either *SQUA* or *FLO* result in a more complete transformation of flowers into inflorescence shoots than do mutations in either *API* or *LFY*. As the expression patterns of *API* and *LFY* differ to some extent from that of their *Antirrhinum* counterparts, perhaps these differences contribute to the different morphologies of *Arabidopsis* and *Antirrhinum* flowers. □

Received 7 August; accepted 28 September 1992.

- Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F. & Meyerowitz, E. M. *Cell* **69**, 843–859 (1992).
- Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. & Sommer, H. *Science* **250**, 931–936 (1990).
- Coen, E. S. & Meyerowitz, E. M. *Nature* **353**, 31–37 (1991).
- Coen, E. S. *et al.* *Cell* **63**, 1311–1322 (1990).
- Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. *Development* **112**, 1–20 (1991).
- Koorneef, M. *et al.* *J. Hered.* **74**, 265–272 (1983).
- Irish, V. F. & Sussex, I. M. *P. Cell* **2**, 741–753 (1990).
- Sommer, H. *et al.* *EMBO J.* **9**, 605–613 (1990).
- Yanofsky, M. F. *et al.* *Nature* **346**, 35–39 (1990).
- Huijser, P. *et al.* *EMBO J.* **11**, 1239–1249 (1992).
- Jack, T., Brockman, L. L. & Meyerowitz, E. M. *Cell* **68**, 683–697 (1992).
- Norman, C., Runswick, M. & Pollock, R. *Cell* **55**, 989–1003 (1988).
- Passmore, S., Maine, G. T., Eble, R., Christ, C. & Tye, B. K. *J. molec. Biol.* **204**, 593–606 (1988).
- Hayes, T. E., Sengupta, P. & Cochran, B. H. *Genes Dev.* **2**, 1713–1722 (1988).
- Mueller, C. G. F. & Nordheim, A. *EMBO J.* **10**, 4219–4229 (1991).

- Schwarz-Sommer, Z. *et al.* *EMBO J.* **11**, 251–263 (1992).
- Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. *Genes Dev.* **5**, 484–495 (1991).
- Chang, C., Bowman, J. L., DeJohn, A. W., Lander, E. S. & Meyerowitz, E. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6856–6860 (1988).
- Steinert, P. M. & Roop, D. R. A. *Rev. Biochem.* **57**, 593–625 (1988).
- Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1731–1735 (1991).
- Crawford, N. M., Campbell, W. H. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8073–8076 (1986).
- Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. *Pl. Cell* **2**, 755–767 (1990).
- Drews, G. N., Bowman, J. L. & Meyerowitz, E. M. *Cell* **65**, 991–1002 (1991).

ACKNOWLEDGEMENTS. We thank S. Kempin for help with the RFLP analyses, D. Weigel for the *Ler* cDNA library; P. Doerner and D. Weigel for advice on the RNA *in situ* hybridizations; S. Elledge and R. Davis for the λ -YES cDNA library; J. Alvarez and D. Smyth for providing the *ap1-2* and *ap1-3* mutants; J. Bowman and J. Alvarez for sharing results before publication; H. Ma for technical assistance; and D. Weigel, J. Bowman, and D. Smyth for critical reading of the manuscript. This work was supported by the NSF and the David and Lucile Packard Foundation (M.F.Y.). C.G.B. and B.S. were supported by NIH training grants. The accession number for the sequence reported in this manuscript is Z16421.

Are vertebrate exons scanned during splice-site selection?

Maho Niwa*, Clinton C. MacDonald†
& Susan M. Berger*

* Verna and Marrs McClean Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

† Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University, New Jersey 08544, USA

PAIRWISE recognition of splice sites as a result of a scanning mechanism is an attractive model to explain the coordination of vertebrate splicing¹. Such a mechanism would predict a polarity-of-site recognition in the scanned unit, but no evidence for a polarity gradient across introns has been found^{2–4}. We have suggested that the exon rather than the intron is the unit of recognition in vertebrates⁵ and that polyadenylation and splicing factors interact during recognition of 3'-terminal exons^{6–8}. Interaction is reflected in maximal rates of *in vitro* polyadenylation. If scanning across the exon is operating during this interaction, then insertion of a 5' splice site should depress polyadenylation. Here we report recognition *in vitro* and *in vivo* of a 5' splice site situated within a 3'-terminal exon, and a concomitant depression of polyadenylation and ultraviolet crosslinking of a polyadenylation factor. Decreased crosslinking was only found when the 3' and 5' splice sites were within 300 nucleotides of each other. These results are consistent with an exon scanning mechanism for splice-site selection.

To investigate the possibility of scanning during splice-site selection, we created precursor RNAs containing duplicated sites within an exon. To circumvent problems associated with discriminating between duplicated sites recognized by the same factors, we used a precursor RNA in which a 5' splice site had been inserted within a 3'-terminal exon (Fig. 1a). Such an exon can be recognized as either an internal exon (by recognition of the inserted 5' splice site), or as a 3' terminal exon (by ignoring the inserted 5' splice site). Figure 1 compares *in vitro* polyadenylation of two-exon (Fig. 1a) and one-exon (Fig. 1b) precursor RNAs in which either a wild-type or mutant 5' splice site has been inserted into the terminal exon. Presence of a wild-type 5' splice site depressed polyadenylation fivefold.

Similar constructs were tested for their ability to polyadenylate *in vivo*. For these experiments, the terminal exon bearing the inserted 5' splice site was positioned as the second exon in a four-exon minigene. No poly(A)⁺ messenger RNA resulting from polyadenylation at the end of exon 2 was detected by reverse transcribed polymerase chain reaction (RT-PCR) (Fig. 1c) or RNase protection analysis (Fig. 1d) when the inserted 5' splice site was wild-type. The dominant RNA produced from this construct was a four-exon RNA in which the first half of exon 2 was spliced to exon 3 using the inserted 5' splice within exon 2. When the inserted 5' splice site was mutant, however, a poly(A)⁺ mRNA containing exons 1 and 2, but not 3 and 4, was detected. A similar result was observed using minigenes containing only the first two exons; in this case, however, the presence of a wild-type 5' splice site caused a null phenotype in which no stable RNA was produced (data not shown). Therefore, insertion of a 5' splice site within a terminal exon has two phenotypes: inhibition of polyadenylation within the same exon, and splicing from the inserted 5' splice site to downstream exons.

To investigate whether insertion of a 5' splice site within a terminal exon could also inhibit association of polyadenylation factors, we used ultraviolet irradiation to crosslink the 64K (*M*_r 64,000) cleavage-stimulation factor CstF⁹. CstF crosslinked well to isolated poly(A) sites and 3' terminal exons (Fig. 2a, lanes 3 and 15). Insertion of a wild-type, but not mutant, 5' splice site caused a decrease in crosslinking of CstF (compare lanes 6 and 9 in Fig. 2a). Therefore insertion of a splicing signal within a terminal exon not only depressed polyadenylation, but also inhibits the normal binding of a specific polyadenylation factor, suggesting that inhibition is occurring at the earliest stages of precursor RNA recognition by the processing machinery. Because association of polyadenylation factors is very rapid *in vitro*, this result suggests that the processing decisions made by these altered exons are early decisions in spliceosome assembly.

Inhibition of crosslinking of CstF was not the result of steric hindrance between factors binding to the 5' splice site and the poly(A) site. When the distance between the inserted 5' splice site and the poly(A) site was increased from 163 to 366 nucleotides, both polyadenylation cleavage (data not shown) and ultraviolet crosslinking of CstF (Fig. 2b: compare lanes 9 and 12) remained low. We also investigated the possibility of steric hindrance by reversing the order of the 5' splice site and the poly(A) site within the terminal exon of the minigene shown in Fig. 1c. In this context, the poly(A) site was recognized most

effectively so that the only RNA detected after transfection of this minigene was polyadenylated within the exon (Fig. 3a and b). We detected no RNA product resulting from usage of the downstream 5' splice site. Therefore a 5' splice site in a terminal exon inhibits polyadenylation and can only be used when upstream of the poly(A) site.

Vertebrate internal exons are rather small, with an average size of 137 nucleotides¹⁰. We have shown that exons larger than 300 nucleotides are inefficiently recognized by splicing factors

even when they are terminated with wild-type splicing signals⁵. This distance restriction suggested that a 5' splice site inserted within a terminal exon would only depress polyadenylation if it was close to the start of the exon. We therefore engineered a wild-type 5' splice site at variable distances downstream of a 3' splice site within a terminal exon and monitored the effect on ultraviolet crosslinking of CstF. We found that insertion of a 5' splice site only prevented crosslinking of CstF when the 5' splice site was near the 3' splice site, with a more distal situation having

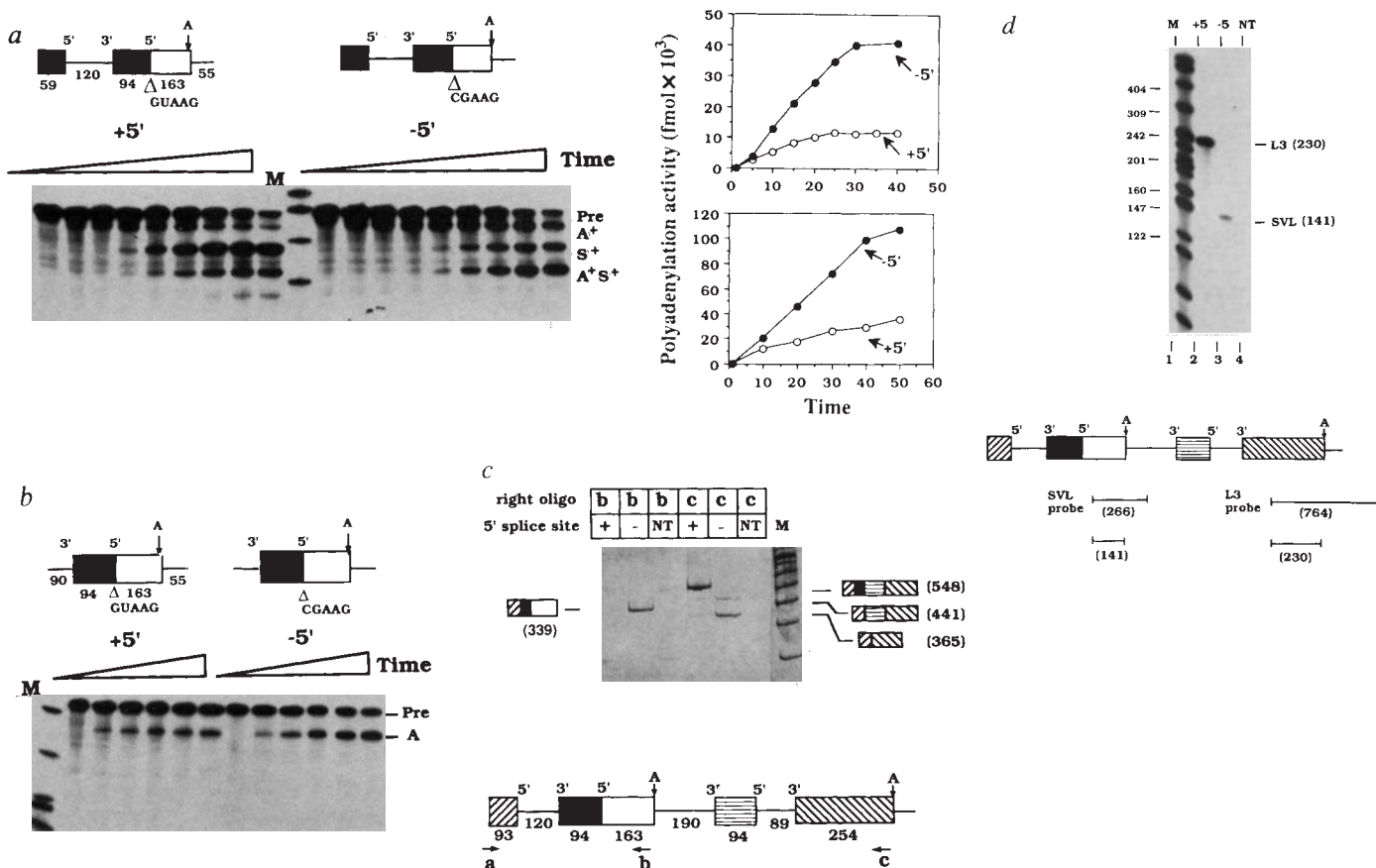
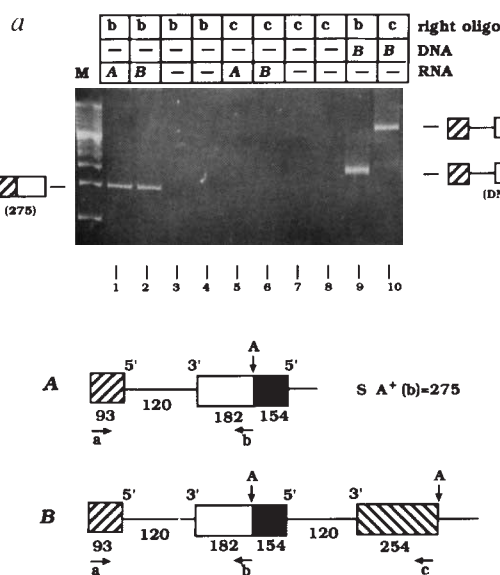


FIG. 1 Insertion of a 5' splice site within a terminal exon depresses *in vitro* and *in vivo* polyadenylation. **a**, **b**, Quantitation of *in vitro* polyadenylation from precursors containing two (**a**) or one (**b**) exons as diagrammed. A precursor for *in vitro* polyadenylation was prepared from a construct in which a 257-nucleotide cassette containing the 5' splice site (plus, wild-type GUAAG; minus, mutant CGAAG) from intron 2 of the major late transcription unit of adenovirus was inserted 41 nucleotides downstream of the 3' splice site in a chimeric terminal exon containing splicing signals from adenovirus and a poly(A) site from SV40 (ref. 11). Product RNAs that had undergone polyadenylation only (A⁺), splicing only (S⁺), or both splicing and polyadenylation (S⁺A⁺) are indicated⁷. Pre, precursor RNA. Polyadenylated product RNA (sum of A⁺ and A⁺S⁺ product in femtomol) are plotted versus reaction time on the right **c**, **d**, Terminal exons from **a** were assayed *in vivo* as part of the minigene shown. All 5' and 3' splice sites are identical and flank exon 2 in the major late adenovirus transcription unit. The poly(A) site (indicated by 'A' in the diagram) terminating exon 2 is from the SV40 late gene and that in exon 4 is from the adenovirus L3 gene. The identities of product RNAs following transfection of HeLa S3 cells were determined by RT-PCR analysis (**c**) using an oligo(dT) reverse-transcription primer and PCR primers as indicated (primers denoted as **a**, **b** and **c**), and by RNase protection analysis (**d**) using a mixture of SV40 and L3-specific RNA probes (SVL, L3). Results were confirmed by northern blot analysis (data not shown). **c**, Lane 1, +5' splice site RNA and exon 2 PCR primer **b**; lane 2, -5' splice site RNA and exon 2 PCR primer **b**; lane 3, non-transfected (NT) cell RNA and exon 2 PCR primer **b**; lane 4, +5' splice site RNA and exon 4 PCR primer **c**; lane 5, -5' splice site RNA and exon 4 PCR primer **c**; lane 6, non-transfected cell RNA and exon 4 PCR primer **c**. The shorter RNAs in lanes 4 and 5 result from RNAs in which neither the 5' splice nor the poly(A) site in exon 2 were recognized and the entire exon was skipped. **d**, Lane 1, *MspI*-digested

PBR322 DNA size markers; lane 2, +5' splice site RNA; lane 3, -5' splice site RNA; lane 4, non-transfected cell RNA. Equal amounts of both indicated probes were used in each analysis. The small amount of L3 protection in the -5' splice site lane presumably arises from the small amount of mRNA produced by complete skipping of exon 2 (see lane 5 of the RT-PCR analysis). The better quantitation afforded by the protection analysis indicated that this event represents less than 5% of the product RNA.

METHODS. *In vitro* processing reactions (conditions permit both splicing and polyadenylation) were done using conditions and precursor RNAs described previously⁷. All reactions contained 3'-dATP to terminate poly(A) addition after a single adenine residue. The amount of polyadenylated RNA was determined by scanning the gel shown in a Betagen Betascope 603 Blot Analyzer. Femtomoles of product RNA were calculated from a knowledge of the specific activity of the input [³²P]UTP-labelled RNA, the U content of each RNA, and the counting efficiency of the instrument. Rates were calculated from the maximal initial slopes. For *in vivo* analysis, total cell RNA from transfected HeLa S3 cells was prepared 48 h after transfection. Minigenes were constructed by replacing the SP6 promoter of the *in vitro* constructs with the Rous sarcoma virus (RSV) promoter. Oligonucleotides were 23-nucleotide sequences from the polylinker beginning exon 1 (primer **a**), the SV40 poly(A) site primer (**b**), and the L3 poly(A) site (primer **c**). The identity of each PCR product band was confirmed by diagnostic digestion with restriction endonuclease. Protection analysis used uniformly ³²P-labelled riboprobes of the indicated sequence. Total cell RNA was hybridized with an equal mixture of the two probes overnight at 45 °C, followed by digestion with a mixture of T1 and T2 ribonucleases. Protected bands are shown on a 10% denaturing polyacrylamide gel. Identity of individual bands was determined by co-migration with *in vitro* processed SV40 and L3 precursor RNAs. Lane M, markers for calibration.

FIG. 3 A 5' splice site within a terminal exon cannot inhibit polyadenylation if it is located downstream of the poly(A) site. The minigenes shown were constructed such that the same 5' splice site and poly(A) site cassettes in the constructs in Fig. 1 were present, but in the opposite order. Product RNAs were analysed using RT-PCR with an oligo(dT) reverse-transcription primer and the PCR primers indicated (a) and by RNase protection analysis using the riboprobes indicated (b). Results were confirmed by analysis of RNA samples by northern blot (data not shown). a, Primers used for PCR were identical to those in the analysis described for Fig. 1c. Lane 1, minigene A RNA and PCR primer b; lane 2, minigene B RNA and PCR primer b; lane 3, non-transfected cell RNA and PCR primer b, lane 4, no RNA; lane 5, minigene A RNA and PCR primer c; lane 6, minigene B RNA and PCR primer c; lane 7, non-transfected cell RNA and PCR primer c; lane 8, no RNA; lane 9, minigene B DNA and PCR primer b; lane 10, minigene B DNA and PCR primer c. Lane M, 100-bp ladder marker. b, Probes used for protection analysis are those described for Fig. 1d. Lane 1, *Msp* I-digested PBR322 DNA size (in nucleotides) markers (M); lane 2, RNA from the transfections using the indicated construct B; lane 3, RNA from non-transfected cells.



METHODS. Total cell RNA was prepared from transfected HeLa S3 cells 48 h after transfection. Product RNA identities were confirmed by diagnostic restriction endonuclease digestion.

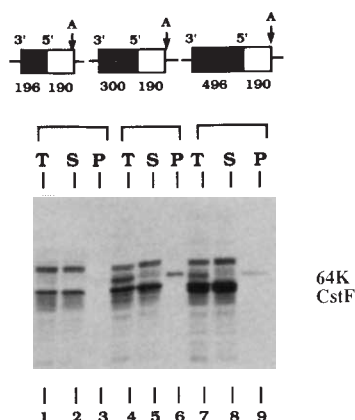
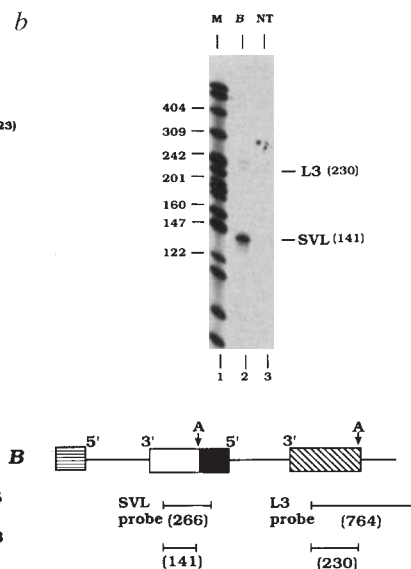


FIG. 4. A 5' splice site inserted within a terminal exon must be close to the 3' splice site to interfere with ultraviolet crosslinking of CstF. Terminal exons in which the distance between the 3' splice site and the inserted 5' splice site varied as shown were prepared and used for *in vitro* crosslinking and subsequent immunoprecipitation of CstF as described for Fig. 2. Quantitation of the immunoprecipitated band indicated at least a fivefold depression of crosslinking in the shortest RNA compared to the longer ones.

METHODS. The diagrammed constructs placed the SV40 poly(A) site various distances downstream of the inserted 5' splice site. The expanded exon sequences are plasmid-derived⁵. Input precursor RNA amounts of 0.50, 0.50, and 0.37 fmol were used for the precursor RNA in lanes 1–3, 4–6 and 7–9, respectively.

little effect (Fig. 4: compare lanes 3, 6 and 9). The inhibition of crosslinking of a polyadenylation factor when a 5' splice site is inserted near the start of a terminal exon indicates that the regions of natural terminal exons proximal to the 3' splice site should be devoid of sequences resembling 5' splice sites. A computer search for possible (cryptic) 5' splice sites in terminal exons revealed that this is indeed the case¹¹.

Our results support the presence of a 5'→3' polarity of site usage across vertebrate exons, consistent with our previous suggestions that the exon is the unit of splice-site recognition in vertebrates. We do not know if an actual scanning mechanism is employed during exon definition. Exons bearing duplicated 5' splice sites do not uniformly show the polarity described here¹²⁻¹⁴, but in this case it is difficult to distinguish between U5-mediated late recognition of 5' splice sites and earlier U1-mediated recognition of the same region^{15,16}.

This study helps to explain earlier observations that poly(A) sites are ignored when they are in vertebrate introns¹⁷⁻¹⁹. Using intact introns, it was impossible to distinguish whether this inhibition resulted from the inability of polyadenylation to compete kinetically with splicing of the intron, or whether the poly(A) site was not recognized because of the interposed 5' splice site. Use of *in vitro* systems and isolated exons showed that the actual binding of polyadenylation factors is inhibited by the presence of an upstream 5' splice site. Combined with the observation that the 5' splice site is only inhibitory to binding if proximal to the start of the exon, these results favour the last interpretation. □

Received 29 May; accepted 25 September 1992.

1. Aeby, M., Hornig, H. & Weissmann, C. *Trends Genet.* **3**, 102-107 (1987).
2. Kuhne, T. *et al.* *EMBO J.* **2**, 727-733 (1983).
3. Lang, K. M. & Spritz, R. A. *Science* **220**, 1351-1355 (1983).
4. Reed, R. & Maniatis, T. *Cell* **46**, 681-690 (1986).
5. Roberson, B. L., Cote, G. J. & Berget, S. M. *Mol. cell. Biol.* **10**, 84-94 (1990).
6. Niwa, M., Rose, S. D. & Berget, S. M. *Genes Dev.* **4**, 1552-1559 (1990).
7. Niwa, M. & Berget, S. M. *Gene Exp.* **1**, 5-14 (1991).
8. Niwa, M. & Berget, S. M. *Genes Dev.* **5**, 2086-2095 (1991).
9. Takagaki, Y. *et al.* *Genes Dev.* **4**, 2112-2120 (1990).
10. Hawkins, J. D. *Nucleic Acids Res.* **16**, 9893-9908 (1988).
11. Brunak, S. *et al.* *J. molec. Biol.* **220**, 49-65 (1991).
12. Eperon, I. P., Estibeiro, J. P. & Eperon, I. C. *Nature* **324**, 280-282 (1986).
13. Cunningham, S. A. *et al.* *J. molec. Biol.* **217**, 265-281 (1991).
14. Lear, A. L. *et al.* *J. molec. Biol.* **211**, 103-115 (1990).
15. Chabot, B. & Steitz, J. A. *Molec. cell. Biol.* **7**, 281-293 (1987).
16. Newman, A. & Norman, C. *Cell* **65**, 115-123 (1991).
17. Adami, G. & Nevins, J. R. *EMBO J.* **7**, 2107-2116 (1988).
18. Levitt, N. *et al.* *Genes Dev.* **3**, 1019-1025 (1989).
19. Brady, H. A. & Wold, W. S. M. *Molec. cell. Biol.* **8**, 3291-3297 (1988).

ACKNOWLEDGEMENTS. We thank T. Cooper and O. Guicherit for advice. This research was funded by grants from the American Cancer Society and Public Health Service to S.M.B.

NMR study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(TG₄T)

Fareed Aboul-ela, Alastair I. H. Murchie
& David M. J. Lilley

Department of Biochemistry, The University, Dundee DD1 4HN, UK

MULTISTRANDED DNA structures based upon guanine association have been proposed to be important in the structure of chromosome telomeres¹ and in immunoglobulin class switching². Nucleic acids containing runs of guanine bases form a number of structures *in vitro*³⁻⁶, including fold-back structures (Fig. 1a)⁷⁻⁹ and parallel-stranded quadruplex structures in DNA^{2,10} and RNA¹¹. The features of fold-back structures have now been determined at high-resolution¹²⁻¹⁴. The different structures are probably based on a tetrad of hydrogen-bonded guanine bases (Fig. 1b), with buffer conditions and sequence effects mediating isomerization

between the different forms^{4,15-18}. Here we use NMR spectroscopy to investigate the solution structure of the complex formed by the hexadeoxynucleotide d(TG₄T) in the presence of sodium ions. We have observed the formation of a parallel-stranded quadruplex containing hydrogen-bonded tetrads of guanine. The parallel-stranded form differs significantly from the fold-back form, with individual nucleotide conformations being closer to those of B-form DNA.

In the ¹H NMR spectrum of d(TG₄T) we observe two sets of six aromatic proton resonances (Fig. 1c), with one form predominating at high temperature and the other at low temperature. This is indicative of a slow exchange between two conformations; the rate of quadruplex formation is known to be slow⁴. We have concluded (see Fig. 1) that the high-temperature form is an unfolded single strand, whereas the low-temperature structure involves the formation of multimers. Because of the number of peaks, the structured form must contain a twofold symmetry if duplex, or fourfold if tetraplex.

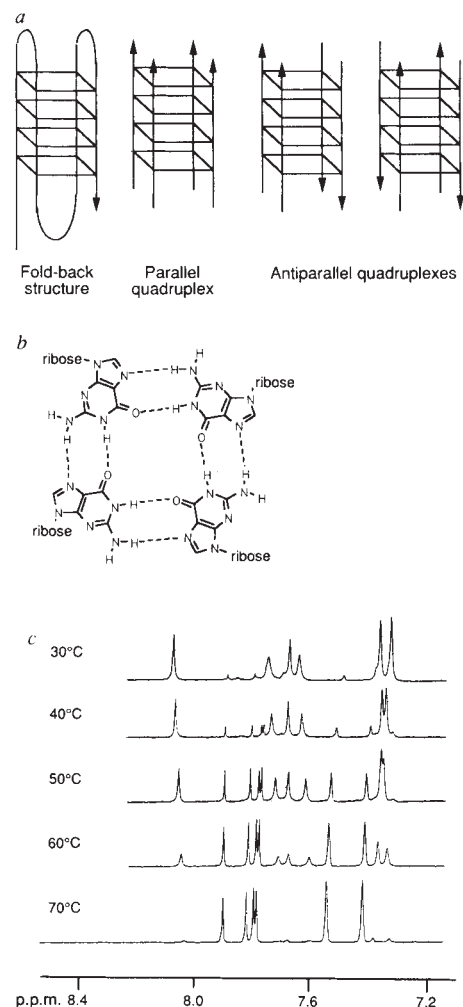


FIG. 1 a, Schematic illustrations of structures proposed to form in G-rich DNA sequences: folded back intermolecular structures, parallel-stranded quadruplex, and antiparallel-stranded quadruplex. b, Guanine tetrad structure, first proposed for 3' GMP fibres²⁴. This arrangement has been observed in fold-back structures by NMR¹⁴ and X-ray crystallography¹³. c, Aromatic spectra of d(TG₄T) in 100 mM sodium chloride, 10 mM sodium phosphate, pH 7.1, 1 mM EDTA in D₂O as a function of temperature. Each spectrum contains twelve aromatic resonances divided into two sets each comprising six resonances, corresponding to two conformations in slow exchange. In addition, four strong and two weak imino resonances (data not shown) disappear together with the broad set of aromatic resonances at high temperature (>70 °C). The narrow lines therefore probably correspond to single-stranded DNA, and the broad lines (7 Hz) indicate the presence of a multimeric, hydrogen-bonded structure at temperatures below 60 °C.

TABLE 1 Parameters describing conformations of individual nucleotides

Residue	$\sum 1' (\pm 0.5)$	$\sum 3' (\pm 1.5)$	S (%)	Ribose pucker	$\chi (\pm 15)$
T1	13.4	10	ND	ND	-140
G2	16	7	>95	C3'- <i>exo</i>	-100
G3	15.3	8	>90	C3'- <i>exo</i>	-140
G4	ND	7	>80	C3'- <i>exo</i>	-150
G5	14.7	8	>80	C3'- <i>exo</i>	-90
T6	14.7	16	60-80	C3'- <i>exo</i>	-140

Predominant sugar conformer and per cent of the time that the individual sugar is in that conformation have been estimated from multiplet widths measured from DQF-COSY (Fig. 2a) according to the method described in Rinkel and Altona²². The glycosidic angle, describing the orientation about the bond connecting the base to the sugar has been estimated by integrating intranucleotide cross peaks from NOESY data, such as that shown in Fig. 3a at a number of mixing times (50, 100, 200, 300 ms), thus allowing for different degrees of build-up of the through-space interaction as well as second-order spin diffusion effects. The results were then compared to those obtained from simulations using the program NOEMOL²³ for the full range of glycosidic angles assuming a south sugar conformation. $\sum 1'$ and $\sum 3'$ are the multiplet widths of the H1' and H3' proton resonances respectively, S is the proportion of south conformer, and χ is the glycosidic torsion angle.

In the presence of 100 mM K⁺ and 10 mM Na⁺ ions at 38 °C, only the six aromatic resonances with identical chemical shift to those of the low-temperature form in 100 mM Na⁺ alone were observed (data not shown).

We felt that it was inadvisable to use peak assignment pathways previously established for B-form DNA^{19,20} as those methods rely on prior assumptions regarding the structure. Instead we used through-bond couplings (data not shown) to assign proton resonances within each deoxyribose, and ¹H-³¹P

couplings (Fig. 2a) to connect them. By this procedure we obtained an unambiguous assignment of the complete backbone (without reference to models), and from these we were able to use nuclear Overhauser effects (NOEs; Fig. 2b) to assign aromatic proton resonances. Assignments of G-imino and G-amino protons involved in base pairing (Fig. 3a) made use of cross-strand NOE interactions whose interpretation was complicated by the symmetry properties of the quadruplex. Therefore the assignment scheme shown in Fig. 3a was tested by observing the spectrum of the hexadeoxynucleotide analogue d(UG₄T) (data not shown). As expected, the imino proton resonance assigned to G2 showed a significant change in chemical shift; the remaining G-imino resonances exhibited much smaller shifts.

The observation of an unbroken path of G H8-ribose H1'-G H8 NOE connectivities along the DNA (in marked contrast to those observed in the fold-back structure¹⁴) suggests that the backbone conformation is similar to that of normal duplex DNA. The relative intensities of NOEs observed between G H8 and ribose H2' compared with H1' indicates that all glycosidic torsion angles are *anti*, and the polarity of the connectivities (G H8 to the ribose protons on the 5' side only) is indicative of a right-handed helix.

As shown in Fig. 3a, imino and amino resonances show unusual chemical shifts and NOE connectivities. Broad peaks were observed in these regions in d5'GMP at 7.2 °C, although no assignment was reported for the peaks near 10 p.p.m.²¹. We observe non-hydrogen-bonded amino protons at 5.6-5.8 p.p.m., with NOE connectivities to the imino and the hydrogen-bonded amino protons (data not shown). These observations indicate that the rotational rate about the C2-N bond in the tetraplex is slow on the NMR timescale.

The fact that individual G imino proton resonances show NOEs to two adjacent aromatic resonances (for example, the

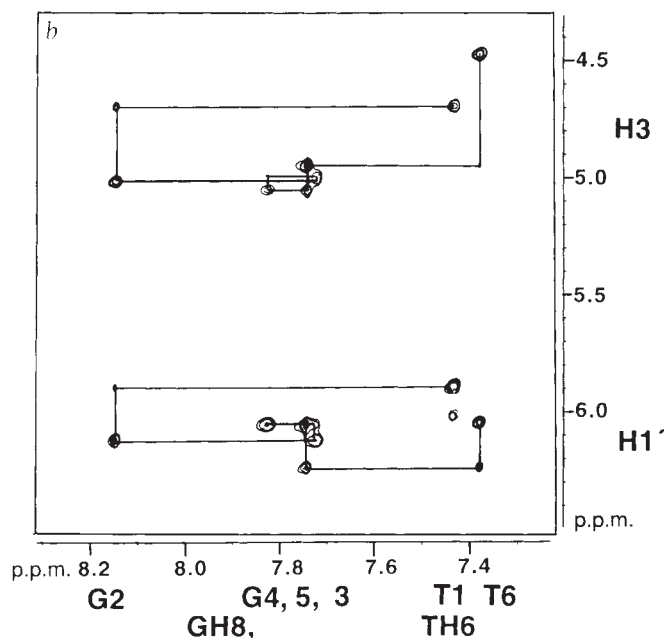
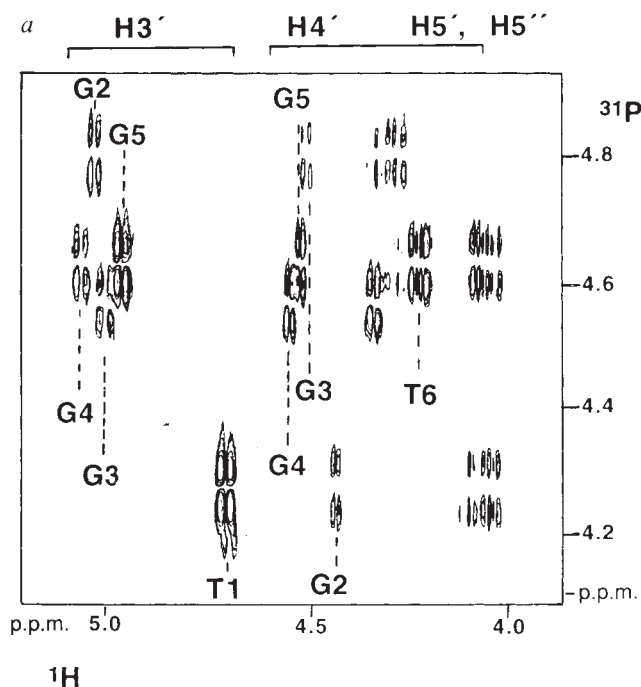


FIG. 2 a, Assignments of backbone resonances using through-bond coupling. Proton-detected heteronuclear correlation²⁵ spectrum of d(TG₄T) in the same buffer and temperature conditions as above. This experiment provides unambiguous identification of each group of sugar resonances determined from the COSY (correlation spectroscopy) experiment (not shown) through the correlation between each of the five phosphorus resonances and the H3' proton of the sugar attached to the 5' neighbouring residue and to each H4' of the sugar attached to the 3' neighbour. 128 data sets containing 2,048 complex data points (128 2K data sets) were acquired with a sweep width of 1,200 Hz in the ¹H dimension, and 160 Hz in the ³¹P dimension. b, Pattern of through-space interactions confirms the presence of right-handed

parallel strands. Nuclear Overhauser enhancement spectroscopy was done under conditions identical to those of COSY experiments. The section of the 300 ms NOESY containing aromatic (GH8 and TH6) resonances on the horizontal axis, and H1', H3' sugar protons on the vertical axis is shown. Sugar protons have been assigned unambiguously from J-coupling experiments. A sequential connectivity pathway is observed between aromatic and H1' protons, and between aromatic and H2', H2'' protons (data not shown) as expected for right-handed helical DNA. Some aromatic to H3' cross-peaks are not present at shorter mixing times, and are probably caused by spin diffusion. 512 2K data sets were acquired with a sweep width of 4,237 Hz.

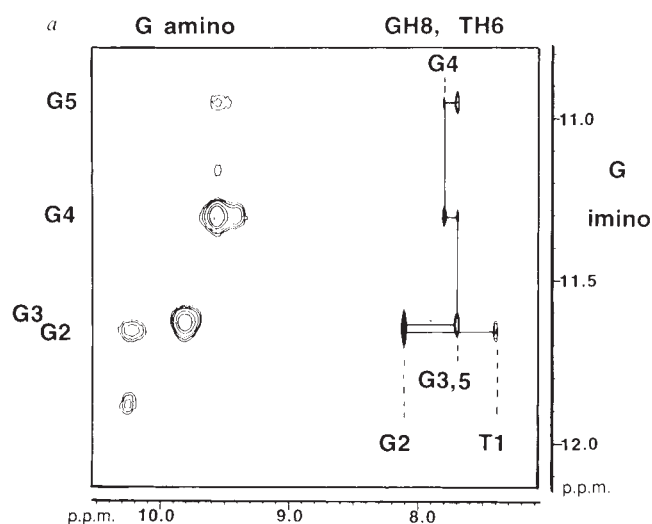


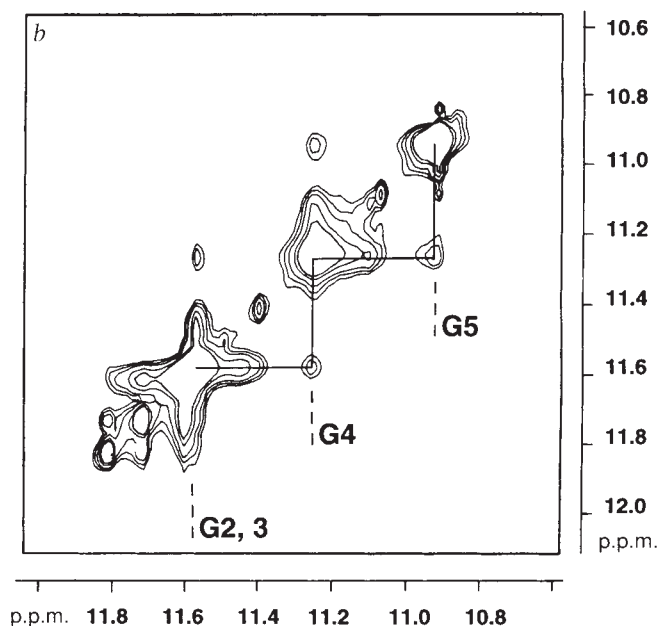
FIG. 3 *a*, NOE spectrum in H₂O of d(TG₄T) in buffer conditions identical to those described in Fig. 1 legend, but at a temperature of 6 °C. Assignments of exchangeable protons are shown. In previous observations of duplex DNA, guanine amino protons have resonated further upfield^{26,27}. Guanine imino protons appear further upfield from their usual position in Watson-Crick G·C base pairs in a chemical shift region (10.5–12 p.p.m.) similar to that observed for G imino protons hydrogen-bonded to carbonyl groups^{3,12,14,16,17,28}, but also for those that are not hydrogen-bonded^{29,30}. These unusual chemical shifts for imino and amino protons are indicative of novel hydrogen bonding. Spin diffusion is the most likely mechanism for the appearance of the intranucleotide imino-GH8 crosspeaks, according to the base tetrad scheme (Fig. 1*b*). If these strands were antiparallel spin diffusion crosspeaks would be expected to be visible between G2H1 and G5H8, and vice versa. 512 K data sets were acquired using time proportional

imino at 11.6 p.p.m. has NOEs to T1 H6 and G2 H8) indicates that at least one of these NOEs is likely to be an intra-base-pair or intra-tetrad contact. This is consistent with the hydrogen-bonding scheme of the tetrad (Fig. 1*b*), and all other base associations that we have been able to devise are not consistent with all the data. From the number of resonances observed we conclude that d(TG₄T) associates as a tetraplex with fourfold symmetry.

Because the parallel-stranded structure is fourfold symmetric, imino-imino contacts within the tetrad will be between magnetically equivalent protons, which will not lead to observable NOEs. In contrast, an antiparallel structure will lead to two sets of intra-tetrad NOEs (associating G3 with G4, and G2 with G5). In the nuclear Overhauser enhancement spectrum recorded in H₂O we observe a sequential pattern of NOEs (Fig. 3*b*), probably arising from inter-base-pair or inter-tetrad contacts, rather than the associations predicted for an antiparallel structure. Moreover, although we observed apparent intra-tetrad imino-aromatic spin diffusion NOEs (Fig. 3*a*) at 6 °C, these did not involve the associations predicted for an antiparallel structure. Thus we find that the structure is a parallel-stranded tetraplex. This conclusion emerges from the NMR data alone, but is fully consistent with available biochemical data.

Sugar conformations and glycosidic rotation angles have been estimated as described in the legend to Table 1. Unlike the fold-back structure in which a pattern of alternating *anti*-*syn* glycosidic angles has been observed, we find all the nucleotides in the parallel-stranded structure are in the *anti* conformation associated with A- or B-form DNA. This suggests that *syn* guanines in the fold-back structure are linked to the steric constraints introduced within antiparallel strands. Sugar conformation is also very similar to that observed in B-form rather than A-form DNA.

In conclusion, we have found that d(TG₄T) forms a parallel-stranded tetraplex with guanines involved in hydrogen-bonded tetrads. This demonstrates that the most stable G₄ tetrad struc-



ture involves parallel (rather than antiparallel) association of strands, in the absence of the entropic advantage inherent in intramolecular folding which may force an antiparallel association. The parallel structure differs substantially from the fold-back structure at the individual nucleotide level. The structural polymorphism of these sequences may be biologically important.

Received 1 June; accepted 1 October 1992.

- Blackburn, E. H. *Nature* **350**, 569–573 (1991).
- Sen, D. & Gilbert, W. *Nature* **334**, 364–366 (1988).
- Henderson, E., Hardin, C. C., Wolk, S. K., Tinoco, I. Jr & Blackburn, E. *Cell* **51**, 899–908 (1987).
- Oka, Y. & Thomas, C. A. *Nucleic Acids Res.* **15**, 8877–8898 (1987).
- Guschlbauer, W., Chantot, J.-F. & Thiele, D. *J. biomolec. struc. Dynam.* **8**, 491–512 (1990).
- Sundquist, W. I. in *Nucleic Acids and Molecular Biology* (eds Eckstein, F. & Lilley, D. M. J.) 1–24 (Springer, Berlin and Heidelberg, 1991).
- Williamson, J. R., Raghuraman, M. K. & Cech, T. R. *Cell* **59**, 871–880 (1989).
- Henderson, E., Moore, M. & Malcolm, B. A. *Biochemistry* **29**, 732–737 (1990).
- Sundquist, W. I. & Klug, A. *Nature* **342**, 825–829 (1989).
- Murchie, A. I. H. & Lilley, D. M. J. *Nucleic Acids Res.* **20**, 49–53 (1992).
- Kim, J., Cheong, C. & Moore, P. B. *Nature* **351**, 331–332 (1991).
- Wang, Y. et al. *J. molec. Biol.* **222**, 819–832 (1991).
- Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. *Nature* **356**, 126–131 (1992).
- Smith, F. W. & Feigon, J. *Nature* **356**, 164–168 (1992).
- Sen, D. & Gilbert, W. *Nature* **344**, 410–414 (1990).
- Hardin, C. C., Henderson, E., Watson, T. & Prosser, J. K. *Biochemistry* **30**, 4460–4472 (1991).
- Hardin, C. C., Watson, T., Corrigan, M. & Bailey, C. *Biochemistry* **31**, 833–841 (1992).
- Lu, M., Guo, Q. & Kallenbach, N. R. *Biochemistry* **31**, 2455–2459 (1992).
- Hare, D. R. et al. *J. molec. Biol.* **171**, 319–336 (1983).
- Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
- Pinnavaia, T. J., Miles, H. T. & Becker, E. D. *J. Am. chem. Soc.* **97**, 7198–7200 (1975).
- Rinkel, L. J. & Altana, C. *J. biomolec. struc. Dynam.* **4**, 621–644 (1987).
- Forster, M., Jones, C. & Mulloy, B. *J. molec. Graph.* **7**, 196–201 (1989).
- Gellert, M., Lipsett, M. N. & Davies, D. R. P. *Proc. natn. Acad. Sci. U.S.A.* **48**, 2013–2018 (1962).
- Sklenar, V., Miyashiro, H., Zon, G., Miles, H. T. & Bax, A. *FEBS Lett.* **208**, 94–98 (1986).
- Boelens, R., Scheek, R. M., Dijkstra, K. & Kaptein, R. *J. mag. Res.* **62**, 378–386 (1985).
- Fazakerly, G. V., van der Marel, G. A., van Boom, J. H. & Guschlbauer, W. *Nucleic Acids Res.* **12**, 8269–8279 (1984).
- Oda, Y., Uesugi, S., Ikehara, M., Kawase, Y. & Ohtsuka, E. *Nucleic Acids Res.* **19**, 5263–5267 (1991).
- Li, Y., Zon, G. & Wilson, W. D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 26–30 (1991).
- Lane, A. N., Jenkins, T. C., Brown, D. J. & Brown, T. *Biochem. J.* **279**, 269–281 (1991).
- Ernst, R. R., Bodenhausen, G. & Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Clarendon, Oxford, 1987).

ACKNOWLEDGEMENTS. We thank D. Patel, S. Homans, C. W. Hilbers and H. Heuss for discussion, and the SERC and CRC for financial support.